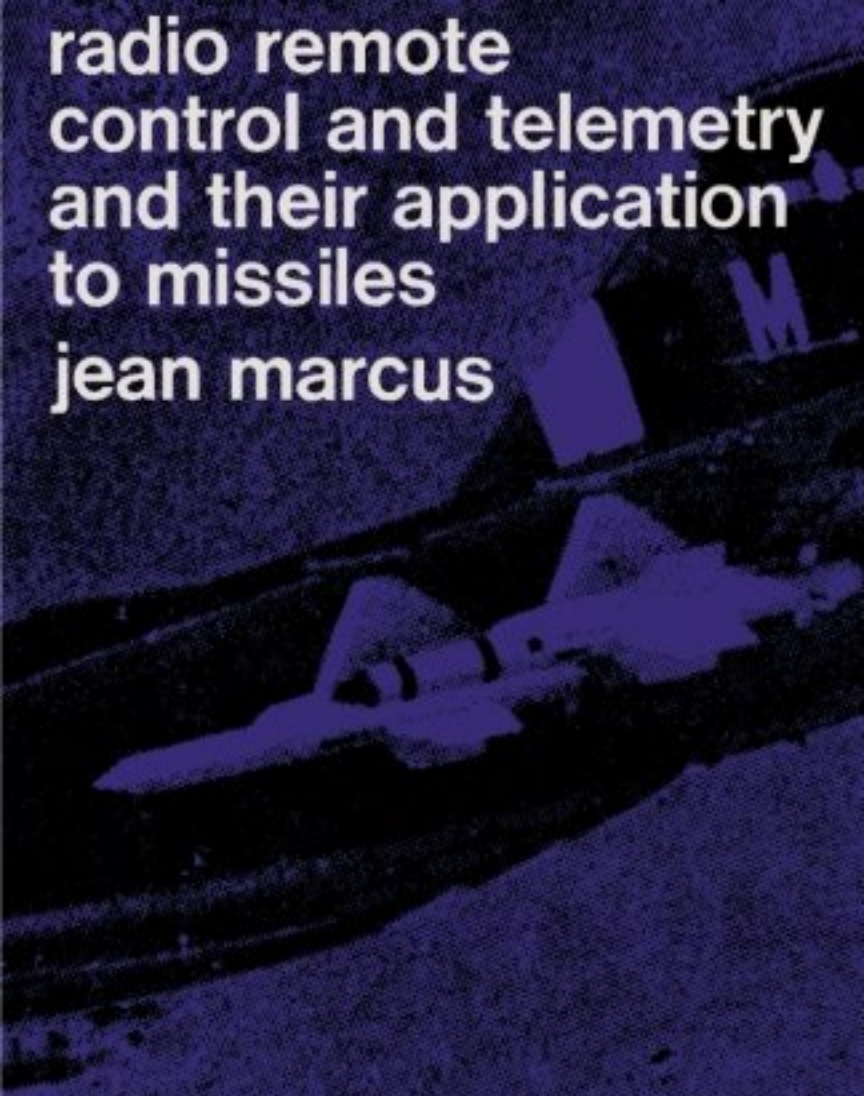
radio remote control and telemetry and their application to missiles

jean marcus



pergamon press

Radio Remote-Control and Telemetry and their Application to Missiles

BY

JEAN MARCUS

Ingénieur de
l'Ecole Nationale Supérieure Electrotechnique,
d'Electronique et d'Hydraulique de Toulouse
Licencié ès Sciences

WITH A PREFACE BY

EMILE STAUFF

Ingénieur en Chef de l'Air
Directeur des Engins Spéciaux à Nord-Aviation

PERGAMON PRESS

OXFORD • LONDON • EDINBURGH • NEW YORK
TORONTO • PARIS • FRANKFURT

Pergamon Press Ltd., Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W. 1

Pergamon Press (Scotland) Ltd., 2 & 3 Teviot Place, Edinburgh 1

Pergamon Press Inc., 44-01 21st Street, Long Island City, New York 11101

Pergamon of Canada, Ltd., 6 Adelaide Street East, Toronto, Ontario

Pergamon Press S.A.R.L., 24 rue des Écoles, Paris 5e

Pergamon Press GmbH, Kaiserstrasse 75, Frankfurt-am-Main

Copyright © 1966
Pergamon Press Ltd

First English edition 1966

Library of Congress Catalog Card No. 65-14225

This is a translation of the orginal French book
Télécommandes et télémesures radio appliquées aux engins spéciaux published by Editions Eyrolles, Paris.

Translated by P. W. HAWKES

1261/66

Copyrighted Material

CONTENTS

PREFACE xi

CHAPTER I. GENERAL REMARKS 1

1. The different types of special missile 2
2. The missile guidance problem 3
3. General remarks about the requirements of the electrical equipment 4
4. The sources of error and their effects 6
5. Limitations of weight and bulk 7

CHAPTER II. THE DIFFERENT TYPES OF MODULATION 8

1. Amplitude modulation (AM) 8
   1.1. Analytic expressions and general characteristics 8
      1.1.1. Analytic expressions 8
      1.1.2. The power of the modulated wave 9
   1.2. The spectrum of an AM wave 10
   1.3. The production of AM waves 11
      1.3.1. Modulation by the grid or by the base 11
      1.3.2. Modulation by the anode or by the collector 12
   1.4. Detection of an AM wave 13
   1.5. Some remarks on AM modulation with a single sideband (SSB) 15
      1.5.1. The principle of SSB modulation 15

2. Frequency modulation and phase modulation 17
   2.1. Analytic expressions and general features 17
      2.1.1. Analytic expressions 17
      2.1.2. The spectrum of an FM wave, and the bandwidth 18
      2.1.3. The power carried by the sidebands 19
      2.1.4. Modulation by a rectangular signal 19
   2.2. The production of frequency-modulated waves and phase-modulated waves 20
      2.2.1. Production of an FM wave by mechanical variations of an oscillator 20
      2.2.2. Devices using variable reactance valves or transistors 22
      2.2.3. Phase modulation devices 23
   2.3. Reception of FM and PM waves 25
      2.3.1. Limiters 25
      2.3.2. Discriminators 28

Copyrighted Material

Copyrighted Material

3. Pulse modulation 29
3.1. The spectrum of a train of pulses. The case of modulated pulses 29
3.2. Bandwidth 33
3.3. Modulation of the pulses 34
3.3.1. Pulse amplitude modulation (PAM) 35
3.3.2. Modulation of the pulse duration (PDM) 35
3.3.3. Pulse position modulation (PPM) 37
3.3.4. Detection of the modulated pulses 37

4. Coded modulation (PCM) 37
4.1. Evaluation of the bandwidth in PCM 37
4.1.1. Bandwidth for a PCM–FM transmission 39
4.1.2. The bandwidth for PCM–AM transmission 40
4.2. Synchronization 40
4.3. Transmission 41
4.4. The receiver 43

Appendix: Gaussian pulses 45

CHAPTER III. INFORMATION AND CODING 47

1. The nature of the information 47

2. A measure of the information 48

3. The capacity of a transmission channel 50

4. Applications to the case of multiplex links 53

5. Some examples of the application of the concept of information efficiency 55

6. Sensible use of information, for long-range links 58
6.1. Limitations on the information to be transmitted 58
6.2. Instruments with a high redundancy 59
6.3. An example of detection with correlation 59

7. Coding 61
7.1. Physical coding 61
7.2. Digital coding 62
7.3. Statistical coding 64
7.4. Secret coding 66

8. Methods of coding and decoding 67
8.1. Digital coding 67
8.2. The parallel–series code converter 70
8.3. Series–parallel code converter 73
8.4. Digital–analogue decoding 74

Appendix: Reasons for choosing the binary system 75

CHAPTER IV. LIMITATIONS ON INFORMATION 77

1. The phenomenon of noise 77

Copyrighted Material

Copyrighted Material

1.1. Noise inherent in the components employed 77
1.1.1. The Johnson effect 77
1.1.2. The shot effect 78
1.1.3. Noise in semiconductors 79
1.1.4. The noise factor 80
1.2. Atmospheric noise 81
1.2.1. Environmental noise 81
1.2.2. Cosmic noise 83
1.3. Pulse noise 84
1.4. Calculation of the signal-to-noise ration for various different types of modulation 85
1.4.1. The signal-to-noise ratio in SSB modulation 85
1.4.2. The signal-to-noise ratio in FM modulation 86
1.4.3. The signal-to-noise ratio in pulsed modulation (PAM) 86
1.4.4. Signal-to-noise ratio in PPM and PDM modulation 87
1.4.5. The signal-to-noise ratio in coded modulation (PCM) 88

2. Interference and jamming phenomena 89
2.1. General introduction 89
2.2. Interference phenomena 90
2.3. Jamming 92

3. Errors and thresholds of measurement 94
3.1. General points 94
3.2. Static errors 96
3.3. Dynamic errors 98
3.3.1. The effect of acceleration on the behaviour of radio components 98
3.3.2. Characteristics of the error produced by acceleration 99
3.3.3. Errors due to temperature effects 102

4. Distortion due to non-linearity 105
4.1. Amplitude distortion 105
The distortion produced by an FM discriminator 106
4.2. Phase distortion 108
4.2.1. Variation of phase delay in amplifiers and filters 108
The influence of phase delay on multiplier stages 110
Phase distortion at the output of a frequency discriminator 110
4.2.2. Variation of phase delay in transmission lines 111
4.2.3. Variation of the phase delay by reflection 112

CHAPTER V. FIELD PROPAGATION AND AERIALS ON MISSILES 115

1. General remarks 115
1.1. The gain, the effective area of an aerial, and the attenuation during propagation 115

2. Low altitude missiles 116
2.1. Reflection from the ground 117
2.2. Tropospheric propagation 119

3. Missiles travelling at great altitudes 120

Copyrighted Material

Copyrighted Material

4. Variations in the gain of aerials 123

5. The problem of attaching aerials to a missile 123
5.1. Classification of aerials and their electrical parameters 124
5.2. The influence of the supporting structure on aerial performance 127
5.3. Description of some types of missile aerials 129
5.3.1. The spiral aerial 129
5.3.2. The slot aerial 131
5.3.3. A special aerial arrangement for use with spinning missiles 132

6. Procedures which are employed to increase the equivalent area of missiles statically 135
6.1. Three mutually perpendicular surfaces 135
6.2. The Luneberg sphere 138

7. Links with diversity 139

CHAPTER VI. RADIO LINKS FOR REMOTE-CONTROL AND TELEMETRY EQUIPMENT, REGARDED AS MULTIPLEX SYSTEMS 143

1. Multiplexing by distribution over frequency 143
1.1. Intermodulation between channels 143
1.2. Selection of channels for remote-control and telemetry equipment with several subcarrier waves 145
1.3. The minimum bandwidth 148
1.4. Signal-to-noise ratios 149

2. Multiplex systems with temporal distribution 151
2.1. Cross-modulation 152
2.1.1. Low frequency cross-modulation 152
2.1.2. Trailing cross-modulation 155
2.1.3. Various other types of cross-modulation 156
2.1.4. The consequences of intermodulation; choice of pulse width 156
2.2. The signal-to-noise ratio 157
2.3. Sampling and distribution procedures 160
2.4. Synchronization 163

3. The threshold range of a radio link 164
3.1. Minimum power at the receiver to ensure adequate protection against noise 165
3.2. Minimum power at the receiver to ensure a given signal-to-noise ratio in the LF channels 166
3.3. The maximum range of the link 167

4. Comparison of different types of modulation and multiplexing 169

CHAPTER VII. THE GUIDANCE OF MISSILES BY RADIO REMOTE-CONTROL 174

1. The position of remote-control in the navigational system 174

2. Piloting by remote-control 176

Copyrighted Material

Copyrighted Material

2.1. The automatic pilot 177
2.2. The remote-control commands and their channel allocation 178
3. Decoding the remote-control commands 180
3.1. The servomotor that actuates the height control 180
3.2. Description of the control equipment in the Nord-Aviation target missile CT 20 182
Stabilization against rolling 182
Stabilization against yawing 182
Stabilization against pitching 183
3.3. Control by mark-space variation 184
Application to flapping aileron control 185
Application to control of the jet deflectors 185
Control by pulse counting 186
4. Examples of radio remote-control systems 187
4.1. An example of FM–FM remote-control 187
4.1.1. General remarks 187
4.1.2. Devices for modulating the subcarrier waves 189
4.1.3. Modulation of the carrier 191
4.1.4. Receiving equipment 196
4.2. An example of digital remote-control 197
4.2.1. General remarks 197
4.2.2. Transmission 197
4.2.3. The receiver unit 198
5. Optimum filtering for missile guidance 199
5.1. Optimization of a linear system 200
5.2. Self-optimization 201
5.3. Combined filtering—application to mixed guidance, using velocity and acceleration 203

CHAPTER VIII. RADIO TELEMETRY EQUIPMENT 206

1. General remarks 206
1.1. The field of application 206
1.2. Quantities which are measured by telemetry 207
1.3. Criteria for a telemetry measurement 207
1.4. FM–FM or FM–AM telemetry 208
1.5. Telemetry with commutation 208
2. Features peculiar to telemetry: the transducers 209
2.1. Strain gauge 209
2.2. Variable voltage or variable current transducers 212
2.2.1. Piezoelectric transducers 212
2.2.2. Photoelectric transducers 212
2.3. Transducers with variable inductance 214
2.4. Transducers with variable capacitance 216
2.5. Transducers employing potentiometers 218
3. Non-linearity in transducers 219
4. The converters 221
4.1. Direct use of variations of inductance or capacitance 222

Copyrighted Material

Copyrighted Material

4.2. The voltage frequency converter 223
4.3. Commutation of telemetry channels 226
4.4. Decommutation at the receiver 230

5. Some practical examples 232
5.1. A PDM–FM telemetry transmitter 232
5.2. Description of a PCM–FM telemetry device (AN–AKT/14) 232
5.3. AJAX, an example of an FM–PM telemetry device 236

6. Recovery of the data 238
6.1. Recording devices 238
6.2. Magnetic recording 239

CHAPTER IX. CONCLUSION 242

APPENDIX 245

BIBLIOGRAPHY 251

INDEX 255

Copyrighted Material

PREFACE

WE are already indebted to Jean Marcus for an excellent and comprehensive book on Frequency Modulation.* The author, who is an engineer constantly dealing with practical problems, showed in that book an admirable balance between the theoretical aspects of the subject, sometimes of a complex nature, and the problems associated with applications.

This same point of view seems to have been Monsieur Marcus's starting point in writing his new book on *Radio Remote-Control and Telemetry and their Application to Missiles.*

The subject is as spectacular as it is difficult.

The public throughout the whole world has heard with admiration the announcement of such achievements as the photographing of the hidden face of the moon, where the photograph was transmitted back in response to terrestrial commands. Everyone has followed day after day the accounts of the prodigious distances—measured in tens of millions of kilometres—over which an artificial satellite has transmitted the measurements made within it. In short, remote-control and telemetry have become headline news.

The subject is, however, a difficult one. The "missiles" in question are usually military. Chapters such as those concerned with coding and jamming have been necessarily limited by restrictions on publication.

For the same reasons, many interesting examples of the application of these techniques have had to be left unmentioned. Fortunately, the field remains large enough for us to be given a work which is well-documented and which will, it is believed, be useful to an audience of engineers and technicians, giving as it does a wide coverage of the missile field.

One point which clearly emerges from this work—and the author himself emphasizes it in several places—is that there are no optimum and generally valid solutions to problems of the kind treated here. The constraints which the working conditions impose upon the systems usually oblige one to make a compromise, and many of the characteristics may be very far from their theoretical optimum values. Added to this is the fact that there are in general several ways of obtaining some desired characteristic, with results which may in practice be very different. The result is a series of partial compromises, between bandwidth and redundancy, or between transmitter power, sensitivity and the type of aerial, for example.

* *La Modulation de Fréquence, Théorie et Applications Industrielles*—(Editions Eyrolles, Paris, 1960).

The art of the engineer consists to a large extent in making these compromises satisfactorily, by calculation wherever possible, but otherwise by intuition. In any event, it is only by studying the particular problem in hand as thoroughly as possible that one can arrive at a rational choice.

In our opinion, a major virtue of such a book as this is that it gives a very full review of the different aspects of the practical data associated with this kind of problem, together with the basic details of the solution.

E. STAUFF

Ingénieur en Chef de l'Air
Directeur des Engins Spéciaux
à Nord-Aviation

CHAPTER I

GENERAL REMARKS

DURING the last few years, considerable developments have been made in the design of remote-control and telemetry equipment. This expansion is closely linked to the upsurge of interest in missiles and space rockets. As soon as the distance between the object being guided and the controller becomes great, radio is the only possible means of control and this is especially true when the link is between the earth and space.

In the present book, the main problems met with in remote-control and telemetry are brought out, together with various methods which have been used to solve them, in the field of missiles.

It is felt that a recapitulation of the various types of modulation will be valuable, since it is basic to the subject and will enable the reader to follow this account of the subject without difficulty.

The chapter on information and coding is intended as a means of gaining a logical grasp of the phenomena in general. This chapter is followed by a discussion of the limitations on information, and here the physical limits to which remote-control and telemetry devices are subject can be seen more exactly.

In another chapter the problems of propagation and of aerials are treated; these are very important topics when the rocket is intended not only to reach very great distances, but to follow various different flight paths and attitudes.

Armed with this information, we have devoted a chapter to multiplex techniques, applied in particular to remote-control and telemetry, with the object of providing the engineer with a practical technique.

The two concluding chapters deal with devices for remote-control and telemetry and illustrate the methods and principles set out in the preceding chapters.

Before attacking the problems of remote-control and telemetry proper, we ought first to define these subjects, and also, to place them in the larger field of interest which comprises guidance and the techniques associated with missiles.

The first rockets were built in Germany during the Second World War; these were the V 1 and the V2. Since that time, however, the study of missiles has become of considerable interest throughout the world, and the whole electronics industry has expanded and progressed at a remarkable rate as a result.

I.1. THE DIFFERENT TYPES OF SPECIAL MISSILE

As can easily be imagined, radio techniques have to vary according to the different kinds of missile which are employed. Missiles can be classified in any of three ways – they can be distinguished by the ways in which they fly; or separated according to their actual type; or again, they can be classed by the function which they are to fulfil.

In the first case, we should distinguish between stabilized missiles and missiles which turn on their axis. The spinning missile is usually small in dimensions, of short range, and travels through the air spinning about its axis like a rifle bullet. The stabilized missile, on the other hand, settles into a fixed orientation with respect to some reference direction – the direction of the force of gravity, for example – under the action of an automatic pilot.

In the second case, the missile can be thought of conceptually – it is either a tactical and offensive missile or it is a target missile. The latter is a pilotless aircraft, the function of which is to replace an enemy aircraft, during firing practice; alternatively, it may take the place of a reconnaissance aircraft.

In the third way of classifying these missiles, we consider the end for which they have been designed. They may thus be classified as target rockets, ground-to-ground (G.G.), air-to-air (A.A.), ground-to-air (G.A.), air-to-ground (A.G.), ground-to-sea (G.S.), sea-to-sea (S.S.) etc. Thus, for example, a G.G. rocket is launched from a terrestrial base, and should reach a target which also is terrestrial, whereas a G.A. rocket is fired from a terrestrial base but is aimed at an aerial target (an aircraft or another missile). This is the classification usually adopted by the French industry; it has the advantage of placing the designer in immediate contact with reality; it is inconvenient, however, in that a large number of different techniques is included in each class. A "ground-to-air missile" can equally well be stabilized or spinning about its axis, despite the fact that the technical problems are quite different.

We have not yet listed certain types of missile which possess individual technical characteristics. I am thinking of very long-range projectiles, either ballistic or directed by radio; and of interplanetary rockets which have to be catalogued separately according to my classification.

For the electronics engineer, a vital question has clearly to be answered: is the guidance to be secret or not? For the radio engineer, the aerodynamic conditions are very often not of prime importance. Rather, in some cases he will be obliged to decide whether a radio link is to exist at all, whether that is, the rocket is to be guided by radio or an internal programme. Here, jamming will have to be studied, and one must ensure also that the relevant consequences of the distance, the quantity of information and the secrecy of the communication are compatible.

I.2. THE MISSILE GUIDANCE PROBLEM

When the missile is controlled from a control-room, the latter should be thought of as a link in the complete piloting chain.

Remote-control, as its name infers, represents a collection of elements the overall function of which is piloting at a distance. Commands are to be conveyed to the missile so that the latter may proceed in any way desired. The four basic commands are dive, climb, left and right. To these, we might add such others as the command to level out again, to detonate the warhead, to put the cameras in action, or the recording instruments, to make smoke, to set the homing or self-guiding head in action.

These remote-controls always appear within the missile as the operation of a relay, the starting of a motor, or the appearance of a potential or a current in appropriate circuits.

The means of piloting are three-fold: the remote-control may be conveyed along a wire, or contained in a programme, or transmitted by radio.

When remote-control is established along a wire, the control-station and the missile are connected by a conducting wire which is progressively paid out as the missile moves further away. The remote-control instructions are transmitted along the line in the form of coded impulses, or as frequencies. Jamming can be said to be impossible, as there is no radio link between the control-post and the missile.

When the remote-control has been programmed, the missile is fired ballistically, and after a certain time has elapsed, the duration of which has been settled beforehand, the control system begins to act within the missile itself; for example, a magnetic tape on which all the characteristics of the desired flight have been recorded beforehand, may be started and dictate a series of instructions to a pick-up head. These commands are in the form of a voltage or of a series of pulses which are decoded inside the missile. Jamming is again impossible.

Unlike remote-control by means of a wire, where the distances between controller and missile cannot exceed a few kilometres, programmed control imposes no limit on this distance; the accuracy, however, is clearly less. Only the third case, remote-control by radio, remains, and it is with this case that we shall be particularly concerned in this book. An operator transmits the remote-control instructions by radio from a transmitting post; aboard the missile, a receiver decodes the signals into voltages which are applied to the different control units. Jamming is now a problem, and it is the minimizing of its influence which dictates the choice of the kind of multiplex link which is to be used. As we have just shown, remote-control represents the executive power of the guidance chain; we must, too, know how to control the missile. The latter requires tracking, either by radar or by means of a transmitter–responder device; this instrument provides the position coordinates, cartesian

or polar, of the missile and may, for example, trace them out with a pen-recorder. It is now a question of transmitting back to the control-post various parameters of the missiles in flight. An instrument which provides this information is known as a "telemetry instrument" (an instrument which performs measurements at a distance).

These data or more exactly, this information, is transmitted from the missile towards the guidance-post, and in this differs from the remote-control which originates in the control-post. Measurements made aboard the missile can provide values of pressure, temperature, level in the fuel tank and speed, for example. Such data are particularly valuable at the design stage.

Telemetry can equally well give such indispensable information as altitude, distance at which the explosive charge goes off, or again it may tell us how well the remote-control commands are being passed on.

Telemetry and remote-control have numerous points in common and it is for this reason that I have treated the two problems in a single book. The natures of the modulation, multiplex techniques and coding are the same, the influence of jamming is common to both, and the radio links obey the same laws.

The essential differences lie in the fact that the positions of the transmitter and the receiver are reversed, and further, the mechanisms for handling and finally decoding the instructions in the remote-control apparatus are of a different nature from those which collect the measurements in telemetry. Remote-control and telemetry have in fact very special ends which are quite distinct.

I.3. GENERAL REMARKS ABOUT THE REQUIREMENTS OF THE ELECTRICAL EQUIPMENT

In remote-control and telemetry equipment, the transmission is distributed over a well-determined number of channels. To each channel corresponds either a command or a measurement. If these parameters are to be correctly transmitted, the bandwidth allotted to each channel must be sufficient for the signal not to be adversely affected. It is necessary also that the overall bandwidth should not be excessively large, in order to employ the frequencies involved to the best advantage, and to reduce the possibilities of jamming and interference to the minimum. This delicate problem will be studied in detail in the chapter devoted to multiplex systems. It is very clear that the type of multiplex system chosen must be suited to the information which is to be transmitted. The bandwidth must be relatively large if a fluctuating measurement is to be transmitted, as this is a rapid physical phenomenon; if a command is to be transmitted in yes-no form a long response-time can be tolerated and a much narrower bandwidth is adequate.

If the transmission of information is to be secret, it must be coded in a fairly complicated way, and this implies an increase of bandwidth. We shall

see how these problems can be solved, either by frequency or by time multiplex techniques.

Suppose for the moment that the bandwidth and type of multiplex system have been chosen, the next stage is the choice of operating frequency range.

The value of the carrier frequency is affected by several factors: the bandwidth which is necessary for the transmission of information, the dimensions of the missile and the aerials, the maximum distance from the control-post and the medium through which the missile travels.

The carrier frequency should in general be at least ten times higher than the maximum R.F. bandwidth, as otherwise, the risk of distortion becomes appreciable. Quality factors of the order of at least fifty are acceptable.

Further, the dimensions of the aerials carried by the missile cannot be settled without reference to the dimensions of the missile itself. It is not easy to employ a four metre aerial on a missile two metres long for aerodynamic reasons. In addition, missiles usually carry several relatively large aerials, and it is important, therefore, to ensure that the aerials are sufficiently independent for possible interactions which might have an unfortunate effect upon the instruments to be reduced to a minimum.

The maximum distance at which the missile is to be controlled, or the maximum range, is one of the vital factors which determine the choice of carrier frequency. This maximum range depends upon the power of the transmitter, the gain of the aerials, the noise in the receivers, and the propagation conditions. If unlimited power were available at the transmitter, the other parameters would be negligible; but this is not at all the case. When the control-station is terrestrial, the problems of bulk and of supply are not major ones, and to a certain extent various ranges of power are at our disposal; but this is a very wasteful solution, since the losses in feeders and couplings increase with frequency. As for the power of the telemetry transmitters, which are situated in the missile itself, virtually nothing can be done. The space allocated to the electrical equipment in a missile is limited, and we are familiar with the result that the power is, roughly speaking, proportional to the volume available; these considerations thus provide a limit. The electromagnetic energy can be concentrated into a tolerably narrow beam, but this requires directive aerials, and the higher the frequency, the more easily are such aerials obtainable.

To ensure an adequate radio link, the power at the receiver must be greater than a certain threshold determined by the noise factor of the receiver. This threshold is low when the carrier frequency is low. For this reason, with a carrier frequency of 100 Mc/s, the noise might be 4 db, whereas for a frequency of 100 Mc/s it would be 8·5 db.

The propagation conditions also play their part in determining the choice of frequency. Attenuation of the transmission is more noticeable at higher frequencies. For preference, we should select higher frequencies for short-range links and low frequencies for longer-range transmissions.

When the missile travels at a great altitude, reflection at the ionosphere has to be taken into account. The losses, by reflection and refraction at the various layers surrounding the earth, are dependent upon the frequency selected.

On the other hand if the missile flies at a low altitude, the nature of the terrain over which it passes must be taken into account. The losses by reflection above the sea or a desert land mass are not the same as those which occur when the country is, for example, hilly.

For a general idea of the frequencies which are usually adopted, a few numerical values are given.

In most applications, for short-range telemetry of large capacity the frequency band bounded by 1365 and 1660 Mc/s is employed; when long-range telemetry is to be performed, either the band 225–260 Mc/s or the band 132–150·8 Mc/s is employed.

I.4. THE SOURCES OF ERROR AND THEIR EFFECTS

The limitations inherent in different systems are of two kinds: random errors and systematic errors.

Among errors of the first type are noise due to fluctuations and pulses, jamming, distortions and intermodulations.

The phenomenon of noise is inherent in the very nature of any circuit (particularly of input circuity), in the frequency bands which are used and in the ambient temperature, and random noise can result from external sources.

The jamming may be systematic or simply intermittent; in either case, we arrive at the same conclusions, namely, that to reduce its influence, the bandwidth of the receiver must be reduced and the information to be transmitted must be coded as thoroughly as possible.

The distortion phenomenon is related to the non-linearity of the circuits, and in particular, to those of the frequency modulators and demodulators; we should distinguish between amplitude and phase distortion.

Among the random errors, there remain only those produced by intermodulation or cross-modulation—I am thinking in particular of the cross-modulation which may arise from possible overloading.

To these random errors must be added the various sorts of systematic error.

In remote-control and telemetry system, we often find ourselves restricted by "threshold". Eloquent enough examples are the minimum triggering current of a relay, the sensitivity of a potentiometer, and the thickness of the trace of a photographic recording. But apart from these quantities which cannot be reduced below a certain limit, there are the limitations of the accuracy of the electronic components, for example a quartz oscillator gives a frequency stable to within one part in 10^5.

Drift is a source of errors which form another category of systematic errors; studying this as a function of time gives a means of estimating the ageing of a component. In every system, it is vital that the drifts produced by variations of temperature, pressure, humidity and atmospheric conditions in general should be appreciated. These drifts are most important when the instruments are used in missiles which pass through a series of successive media in which the climatic conditions are very different.

Finally, among systematic errors which are often omitted, we should remember the influence of the human operator who, at some stage or another, makes an appearance in the guidance chain.

When a system is being designed, we must discover the natural bounds by which we shall be limited. Above all, the order of magnitude of these bounds should never be lost sight of; this will keep our ideas realistic.

I.5. LIMITATIONS OF WEIGHT AND BULK

Inside the missile, considerations of weight and volume are critical ones; they, of course, limit the performances of the electronic systems in the instruments. Building the components on a miniature scale often results in our having to choose high frequencies, and in the development of constructional methods specially adapted to this problem; examples are the technological advances in printed circuit construction, and the introduction of transistors into the units. The ever present restrictions on bulk often oblige one to select some material which may sometimes only be indifferent from the electronic standpoint.

Taking all these factors into consideration, therefore, it is clear that we shall have to regard the instruments to be carried aboard the missile from a very special standpoint. The radio engineer working on the problems with which these missiles confront him, cannot but occasionally show a preference for the simplest solution rather than the solution which seems theoretically perfect.

CHAPTER II

THE DIFFERENT TYPES OF MODULATION

THE ENGINEER must be guided in his choice of modulating system by the goal which he plans to achieve. The reasons for his choice will be dictated by certain characteristic quantities, for example – the range of frequencies involved, the linearity of the modulating and demodulating devices, the power which is available, the ease of construction and the signal-to-noise ratio.

In this Chapter, an overall view is attempted of the different types of modulation which are in regular use, and their characteristics are compared with those of amplitude modulation, which is used as a standard for reference. The question of the signal-to-noise ratios will be considered in the chapter devoted to noise, where it fits in naturally.

II.1. AMPLITUDE MODULATION (AM)

II.1.1. *Analytic Expressions and General Characteristics*

II.1.1.1. *Analytic Expressions*

We consider a pure carrier wave, the analytic expression for which is of the form

$$e^{j(\omega_0 t+\varphi)},$$

in which f_0 is the carrier frequency $\left(f_0 = \frac{\omega_0}{2\pi}\right)$

φ is a phase constant.

A wave which has been modulated in amplitude can be expressed in a general way by a function of the form

$$V(t) = V_0 f(t)\, e^{j(\omega_0 t+\varphi)}$$

in which

$$f(t) = 1 + ms(t).$$

$s(t)$ represents the modulating signal which is to be transmitted. The coefficient m is the modulation index. When the signal is sinusoidal we have

$$s(t) = \cos(\Omega t + \psi)$$

and the modulated carrier wave takes the form

$$V(t) = V_0[1 + m\cos(\Omega t + \psi)]\cos(\omega_0 t + \varphi). \qquad \text{(II.1)}$$

When the modulation index is unity, ($m = 1$), the carrier wave is said to be 100 per cent modulated.

Equation (II.1) can be written in the following way:

$$V(t) = V_0\cos(\omega_0 t + \varphi) + \frac{mV_0}{2} \times$$

$$\times\{\cos[(\omega_0 + \Omega)t + \varphi + \psi] + \cos[(\omega_0 - \Omega)t + \varphi - \psi]\}. \qquad \text{(II.2)}$$

We can thus regard the modulated wave as the sum of three vectors:

OA represents $V_0\cos(\omega_0 t + \varphi)$

AB represents $\frac{mV_0}{2}\cos[(\omega_0 + \Omega)t + \varphi + \psi]$

AC represents $\frac{mV_0}{2}\cos[(\omega_0 - \Omega)t + \varphi - \psi]$.

These vectors are shown in Fig. II.1.

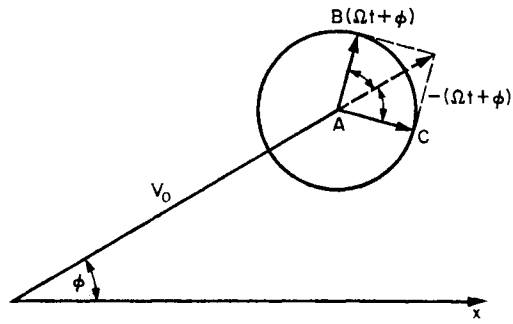


FIG. II.1. Graphical representation of sinusoidal amplitude modulation.

II.1.1.2. *The Power of the Modulated Wave*

If P_0 is the power of the unmodulated carrier wave, the maximum power of the modulated wave becomes

$$P_{max} = (1 + m)^2 P_0. \qquad \text{(II.3)}$$

The mean power is equal to the sum of the power of the unmodulated carrier wave and the mean power contained within the sidebands; it is thus equal to

$$P_{\text{mean}} = P_0 + 2\frac{m^2}{4}P_0,$$

or

$$P_{\text{mean}} = \left(1 + \frac{m^2}{2}\right)P_0. \tag{II.4}$$

II.1.2. *The Spectrum of an AM Wave*

In the most general case, where the signal $s(t)$ is not purely sinusoidal, the modulated carrier wave can be expressed in a Fourier series of the form:

$$V(t) = V_0[1 + a_0 + a_1 \cos(\Omega t + \psi_1) + a_2 \cos(2\Omega t + \psi_2) + \cdots] \cos(\omega_0 t + \varphi),$$

or

$$\frac{V(t)}{V_0} = (1 + a_0)\cos(\omega_0 t + \varphi) + \frac{a_1}{2}[\cos((\omega_0 + \Omega)t + \varphi + \psi_1) + \cos((\omega_0 - \Omega)t + \varphi - \psi_1)] + \frac{a_2}{2}[\cos((\omega_0 + 2\Omega)t + \varphi + \psi_2) + \cos((\omega_0 - 2\Omega)t + \varphi - \psi_2)] + \cdots.$$

The spectrum of the modulated wave consists of a central line, which corresponds to the carrier frequency f_0, together with a series of lines of higher and lower frequencies, symmetrically placed with respect to the central frequency. These lines lie at frequencies

$$f_0 + F, \quad f_0 + 2F, \quad \ldots \text{ etc. } \ldots \quad \text{and} \quad f_0 - F, \quad f_0 - 2F, \quad \ldots \text{ etc. } \ldots$$

In the special case of sinusoidal modulation, the spectrum consists of three lines: the central line, with amplitude V_0, and the two adjacent lines at frequencies $f_0 + F$ and $f_0 - F$, each with amplitudes of $mV_0/2$.

As a result, the width of the frequency bandwidth occupied by an AM wave, modulated by a sinusoidal signal, is given by

$$\mathscr{B} = 2F. \tag{II.5}$$

II.1.3. *The Production of AM Waves*

Two modulation procedures are to be distinguished: modulation by the grid and modulation by the anode.

II.1.3.1. *Modulation by the Grid or by the Base*

An example of modulation by means of the grid is shown in Fig. II.2. The instantaneous voltage at the grid is

$$u = A \sin \Omega t + B \sin \omega t + U_0.$$

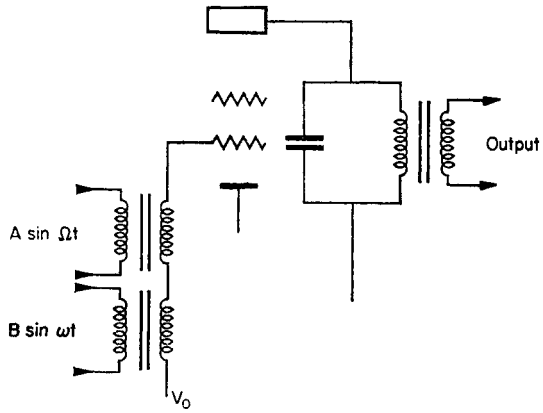


FIG. II.2. Grid AM modulation.

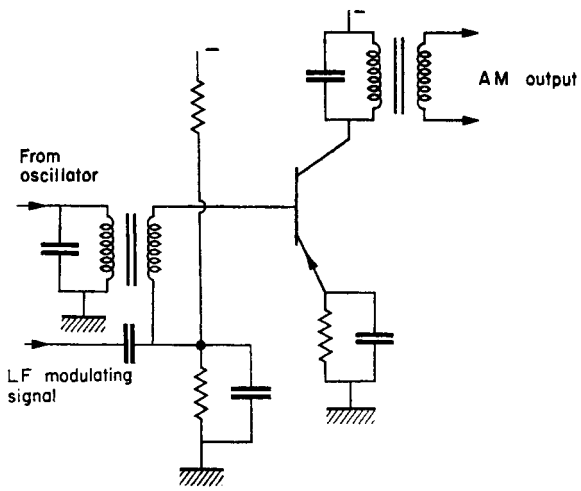


FIG. II.3. Base AM modulation.

During one period of the carrier, the modulating voltage can be considered a constant, equal to $U_0 + A \sin \Omega t$, and it varies with the same period as the LF between $U_0 + A$ and $U_0 - A$. The value of A depends upon the modulation factor which is chosen. In Fig. II.3 an arrangement where the modulation occurs at the base of a transistor is shown.

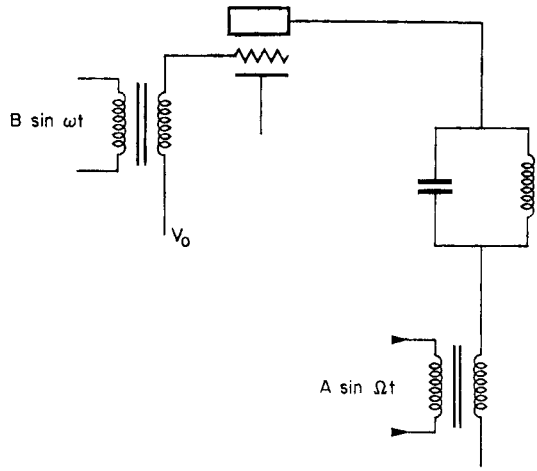


FIG. II.4. Anode AM modulation.

II.1.3.2. *Modulation by the Anode or by the Collector*

In the case the grid voltage is held constant at U_0, the carrier signal is applied to the grid and the LF signal to the anode, as shown in Fig. II.4. The modulated signal appears across the tuned anode load. The methods shown in Fig. II.5 employ capacitors or inductors for modulation, and for

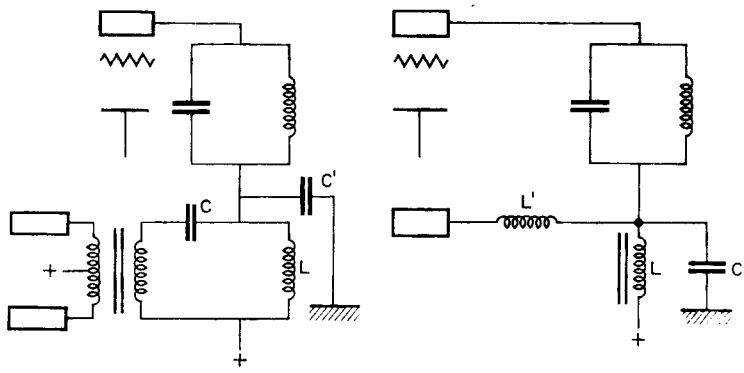


FIG. II.5. The use of capacitors (C) or inductors (L) for modulation.

preference, one of these methods is usually selected, in order to avoid polarization of the modulating transformer.

A type of AM arrangement in which the LF is introduced at the collector of the transistor, is shown in Fig. II.6.

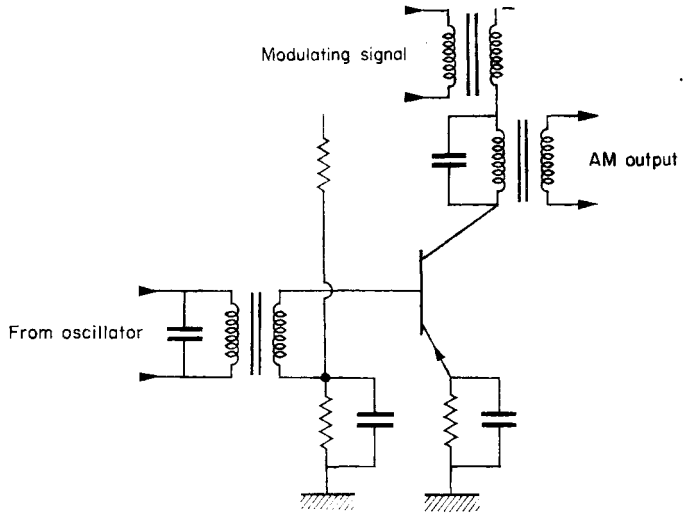


FIG. II.6. Collector AM modulation.

II.1.4. *Detection of an AM Wave*

The detectors which are used in practice are not exactly linear; instead, they are usually "quadratic"—the signal $u(t)$ which is detected is expressed as a function of the modulated wave $V(t)$ by the relation

$$u(t) = k_1[V(t)]^2 + k_2 V(t). \tag{II.6}$$

On substituting the expression for $V(t)$ given in equation (II.1), and considering the case in which both φ and ψ are zero, we have

$$u(t) = k_1 V_0^2[1 + m\cos\Omega t]^2 \cos^2(\omega_0 t) + k_2 V_0[1 + m\cos\Omega t]\cos(\omega_0 t).$$

On replacing the terms in $\cos^2 x$ by $(1 + \cos 2x)/2$ and terms in products of cosine functions by terms in sums, we find

$$u(t) = \frac{k_1 V_0^2}{2}\left(1 + \frac{m^2}{2}\right) + \cos\Omega t[mk_1 V_0^2] + \cos 2\Omega t\left[\frac{m^2}{4} k_1 V_0^2\right]$$

$$+ \cos \omega_0 t \, [k_2 V_0]$$

$$+ \cos (\omega_0 + \Omega)t \left[\frac{m}{2} k_2 V_0\right]$$

$$+ \cos (\omega_0 - \Omega)t \left[\frac{m}{2} k_2 V_0\right]$$

$$+ \cos 2\omega_0 t \left[\left(1 + \frac{m^2}{2}\right) \frac{k_1 V_0^2}{2}\right]$$

$$+ \cos (2\omega_0 + \Omega)t \left[m \frac{k_1 V_0^2}{2}\right]$$

$$+ \cos (2\omega_0 - \Omega)t \left[m \frac{k_1 V_0^2}{2}\right]$$

$$+ \cos 2(\omega_0 + \Omega)t \left[\frac{m^2}{4} k_1 \frac{V_0^2}{2}\right]$$

$$+ \cos 2(\omega_0 - \Omega)t \left[\frac{m^2}{4} k_1 \frac{V_0^2}{2}\right].$$

We can now see that the output signal from such a detector is composed of a number of separate signals, the angular frequencies of which are

$$\Omega, \quad 2\Omega$$

$$\omega_0, \quad \omega_0 \pm \Omega$$

$$2\omega_0, \quad 2\omega_0 \pm \Omega, \quad 2\omega_0 \pm 2\Omega.$$

As some of these angular frequencies are close to or equal to that of the carrier wave, a low-pass filter must be placed at the output of the detector to eliminate them. After filtering, the only components which remain are

$$\cos \Omega t (m k_1 V_0^2) + \cos 2\Omega t \left(\frac{m^2}{4} k_1 V_0^2\right).$$

If a comparatively low modulation index is used, the term in 2Ω becomes negligible in comparison with the term in Ω. (For $m = 30$ per cent, the attenuation of the 2Ω component, with respect to that of the Ω component, is equal to $m/4$ which is less than 10 per cent.)

In Fig. II.7, a crystal detector is illustrated schematically, together with its detection characteristic (the amplitude of the signal at the output as a function of the signal at the imput).

The time constant, RC, is taken to be $1/F_{max}$, where F_{max} is the maximum frequency to be transmitted.

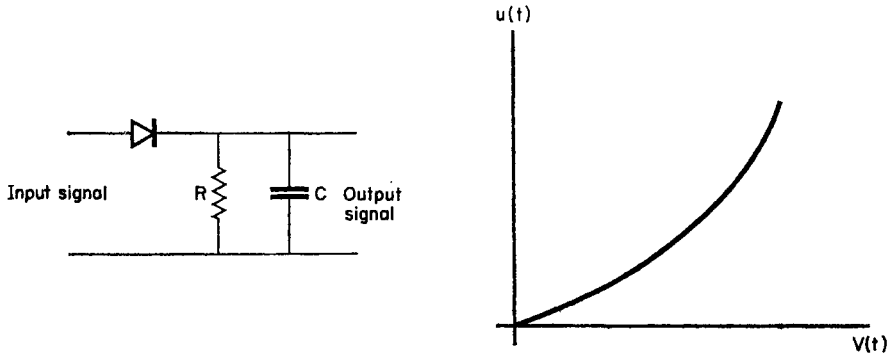


FIG. II.7. Crystal detector.

II.1.5. *Some Remarks on AM Modulation with a Single Sideband (SSB)*

Modulation with a single sideband offers some extremely noteworthy advantages, and this system of modulation is unquestionably among the best-adapted to the kind of information to be transmitted. The value of the signal-to-noise ratio is far higher than the corresponding value with ordinary amplitude modulation. The frequency band which is occupied is only half as wide, and this is an added protection against jamming. The major inconvenience is of a technological nature: it is in fact necessary to have local frequencies available at the transmitter and at the receiver which differ by no more than a few cycles per second. The problems of bulk in the missile are so severe that this type of modulation system is very little used in remote-control and telemetry devices. For this reason that we shall not study it in any detail but simply point out its existence for completeness' sake.

II.1.5.1. *The Principle of SSB Modulation*

The problem is to suppress the component of the spectrum which corresponds to the central frequency, as the uselessness of the latter has already been displayed in the paragraph on AM.

If $V_P(t) = \cos(\omega t + \varphi)$ is the voltage which represents the HF carrier wave, and $s(t) = \sum_{n=1}^{N} E_n \cos(\Omega_n t + \psi_n)$ is the modulation signal, the AM

signal is

$$V(t) = [1 + ks(t)]V_P(t)$$

$$= \cos(\omega t + \varphi) + \frac{k}{2}\sum_{n=1}^{N} E_n \cos[(\omega - \Omega_n)t + \varphi - \psi_n]$$

$$+ \frac{k}{2}\sum_{n=1}^{N} E_n \cos[(\omega + \Omega_n)t + \varphi + \psi_n].$$

Suppose that with some appropriate device, their product is formed instead:

$$U(t) = V_P(t) \times s(t)$$

$$= \tfrac{1}{2}\sum_{n=1}^{N} E_n \cos[(\omega - \Omega_n)t + \varphi - \psi_n]$$

$$+ \tfrac{1}{2}\sum_{n=1}^{N} E_n \cos[(\omega + \Omega_n)t + \varphi + \psi_n].$$

We obtain a wave, whose spectral components consist only of two sidebands, lying on either side of the carrier frequency. To obtain a signal containing a single sideband, therefore, it is only necessary to filter out one of the two bands. When the lower band is filtered out, the SSB signal is of the form

$$V_{SSB}(t) = \tfrac{1}{2}\sum_{n=1}^{N} E_n \cos[(\omega + \Omega_n)t + \varphi + \psi_n].$$

To detect a SSB wave, the SSB signal which is to be demodulated and a demodulation signal

$$V_D(t) = \cos(\omega_D t + \varphi_D)$$

are applied to the demodulation device simultaneously; at the output of the mixer we obtain the signal $V_s(t)$:

$$V_s(t) = \sum_{n=1}^{N} E_n \cos[(\omega - \omega_D + \Omega_n)t + \varphi - \varphi_D + \psi_n].$$

If $\omega_D = \omega$, the output signal $V_s(t)$ is thus the modulation signal $s(t)$, apart from a phase factor, φ_D.

The difficulty is to obtain $\omega_D = \omega$ exactly. If this equality is not exactly achieved, serious perturbations to the LF signal itself may result. The solution is to force the mixer oscillator by means of the error in frequency between the signal, frequency f, and the signal provided by an extremely accurate local oscillator.

An example of a method which can be used to suppress the carrier frequency is given in Fig. II.8. At the output of this balanced modulator, a

signal with a pair of sidebands is obtained; the carrier wave has been suppressed; to eliminate one of the two sidebands, this output signal has simply to be passed through a filter.

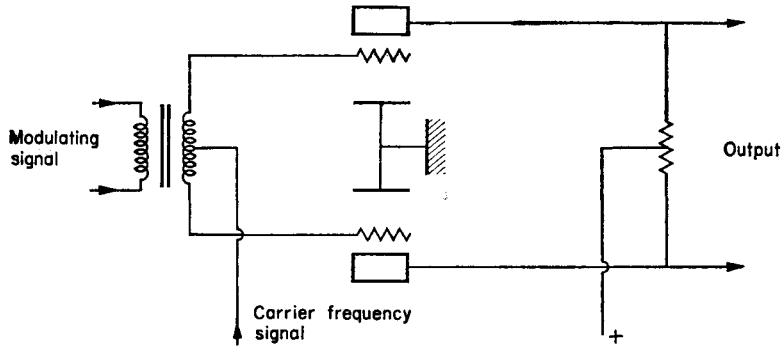


FIG. II.8. Balanced modulator.

II.2. FREQUENCY MODULATION AND PHASE MODULATION[59]

In this section, we shall examine in detail these types of modulation, which are of particular interest in remote-control and telemetry devices.

II.2.1. *Analytic Expressions and General Features*

II.2.1.1. *Analytic Expressions*

Consider a function of the form

$$V(t) = V_0 \sin \varphi(t),$$

in which $\varphi(t)$ is the phase or the argument of the function $V(t)$. By definition, the instantaneous frequency f_i is, apart from a constant 2π, equal to the time derivative of the argument, thus:

$$f_i = \frac{1}{2\pi} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t}.$$

A frequency-modulated wave results if the value of the instantaneous frequency faithfully reproduces that of the modulating signal, $s(t)$.

$$\omega_i = \omega + ks(t).$$

The argument or phase is determined by

$$\varphi(t) = \omega t + k \int s(t)\, \mathrm{d}t + \varphi_0$$

in which ω is the angular frequency of the unmodulated carrier wave.

The general expression for a wave modulated in frequency by a signal $s(t)$ is of the form

$$V(t) = V_0 \sin [\omega t + \varphi_0 + k \int s(t)\, dt].$$

In the special case where $s(t)$ is a circular function, namely

$$s(t) = S_0 \cos \Omega t,$$

we have

$$V(t) = V_0 \sin [\omega t + \varphi_0 + m \sin \Omega t],$$

in which $m = kS_0/\Omega$ is the modulation index.

The instantaneous frequency can be expressed in the form

$$f_i = f + \Delta f \cos \Omega t,$$

$$m = \frac{\Delta f}{F} = \frac{\Delta \omega}{\Omega}.$$

The modulation factor λ is the ratio of the frequency difference Δf, to the carrier frequency f:

$$\lambda = \frac{\Delta f}{f}.$$

In the case of phase modulation,

$$V(t) = V_0 \sin [\omega t + kS_0 \cos \Omega t + \varphi_0],$$

or alternatively

$$V(t) = V_0 \sin \varphi(t)$$

with

$$\varphi(t) = \varphi + \Delta\varphi \cos \Omega t.$$

Writing $S(t) = ds(t)/dt$, a PM wave can be identified as $s(t)$ and an FM wave as $S(t)$.

The transition from one type of modulation to the other is effected simply by differentiation or integration.

II.2.1.2. *The Spectrum of an FM Wave, and the Bandwidth*

It can be shown that an FM wave of the form $V(t) = V_0 \sin (\omega t + m \sin \Omega t)$ can be expanded as an infinite series of the following form:

$$V(t) = V_0 \sum_{n=-\infty}^{\infty} J_n(m) \sin (\omega t + n\Omega)t,$$

in which

$$J_{-n}(m) = (-1)^n J_n(m),$$

and $J_n(m)$ is a Bessel function of order n.

The resulting spectral distribution consists of a series of lines symmetrically placed about the carrier frequency f, and symmetric in amplitude variation.

This means that for even harmonics of the modulating frequency, $(n\Omega)$, the corresponding spectral amplitudes on either side of the carrier frequency are in phase, whereas for odd harmonics, they are of opposite phase. It would seem that the frequency band occupied by the modulation is infinitely wide, since the number of components is unbounded. In fact however, $J_n(m)$ tends rapidly towards zero when n increases, and a practical bandwidth $\mathscr{B}$ can be defined in the following way:

$$\mathscr{B} = 2(\Delta f + F_{max})$$
$$= 2(m + 1)F_{max},$$

in which F_{max} is the maximum LF modulation frequency.

For small values of the modulation index, $(m < 0{\cdot}2)$, the width of the frequency band occupied is equal to twice the modulation frequency (a result analogous to the AM case).

II.2.1.3. *The Power carried by the Sidebands*

If $P_0 = AV_0^2$ is the power of the unmodulated carrier wave, the power contained in all the sidebands of an FM wave is

$$P = P_0[1 - J_0^2(m)].$$

When the modulation index m is greater than 2, the ratio P/P_0 lies between 0·8 and unity, and tends towards unity as m increases; this means that the power carried by all the sidebands tends in a uniform way towards the power in the unmodulated carrier wave. This fact represents an advantage over amplitude modulation, where the power at the aerial of the transmitter depends upon the modulation factor. The result is that firstly, the power supplied to the different amplifying stages is sensibly constant and further, that the useful power is greater than in amplitude modulation.

II.2.1.4. *Modulation by a Rectangular Signal*

This problem appears when it is necessary to modulate a carrier wave with a succession of digits, or again, it appears in a more general way in devices for pulsed modulation which employ frequency modulation of the carrier.

The expression for the signal which has been modulated in frequency can be written in the form

$$V(t) = V_0 \sin [\omega t + \Delta\omega \int s(t)\, dt].$$

By decomposing the rectangular signal into a Fourier series, and then doing the same to the signal $V(t)$, we can show that the carrier frequency F, corresponds to a spectrum consisting of a succession of lines. The height of the nth line is given by

$$A_n = \frac{m}{m+n} \frac{\sin (m-n)\frac{1}{2}\pi}{(m-n)\frac{1}{2}\pi}.$$

The bandwidth which the spectrum occupies remains approximately the same as in the case of modulation by a sinusoidal signal, namely

$$\mathscr{B} = 2(m+1)F.$$

II.2.2. *The Production of Frequency-modulated Waves and Phase-modulated Waves*

Numerous devices have been built which are capable of providing frequency modulation. These can be classified into three broad categories:

- the production of an FM wave by mechanical variation of an oscillator
- the modulation may be obtained by a reactance valve
- phase-modulation devices may be used as intermediaries.

II.2.2.1. *Production of an FM Wave by Mechanical Variations of an Oscillator*

The frequency of an oscillator is determined by the inductance and the capacitance of its resonant circuit ($LC\omega^2 = 1$). To make the frequency of oscillation follow the original variable, therefore, it is only necessary to connect the latter mechanically to vary either the capacitance C or the inductance L.

The relative variations in the frequency are then given by the following relations:

Either

$$\frac{\Delta\omega}{\omega} = \frac{\Delta C}{2C_0}$$

or

$$\frac{\Delta\omega}{\omega} = \frac{\Delta L}{2L_0}.$$

If the changes of capacitance or inductance are sinusoidal in time—if for example

$$C = C_0 + \Delta C \sin \Omega t$$

the instantaneous value of the frequency is given approximately by

$$\omega_i = \omega + \Delta\omega \sin \Omega t.$$

In Fig. II.9, an example of an oscillator with variable capacitance is shown.

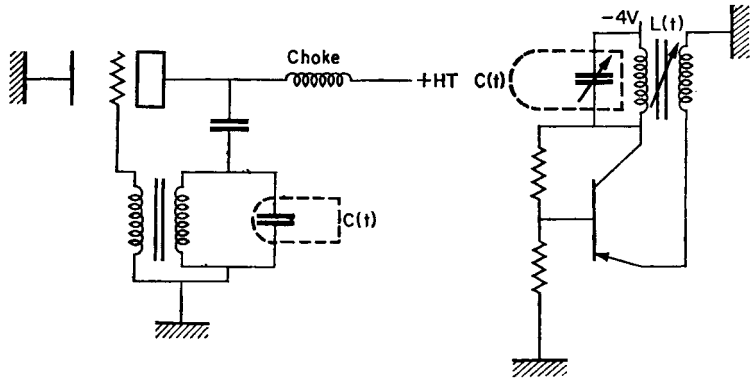


FIG. II.9. Variable-capacitance circuits.

In variable-inductance oscillators, it is generally the permeability of the core which is selected as the variable parameter. This can be effected either by varying the gap or by using a core in the form of a plunger, the position of which is coupled to the quantity which is varying, or again by altering the

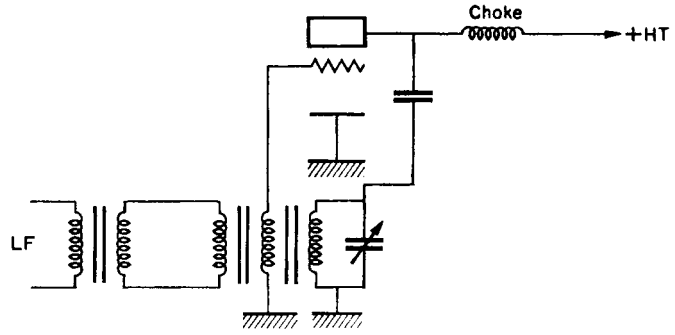


FIG. II.10. Production of FM signal, using magnetic circuits.

saturation of the magnetic material by means of an auxiliary winding (see Fig. II.10).

These arrangements are very often used in telemetry and will be examined in more details in the section devoted to transducers (Chapter VIII).

II.2.2.2. *Devices Using Variable Reactance Valves or Transistors*

Consider the circuit given in Fig. II.11; it can be shown that the impedance of the valve together with the circuits Z_1 and Z_2 is equivalent to an impedance Z_i across the terminals of the oscillator V_e and which is given by

$$Z_i = \frac{Z_1 + Z_2}{g_m Z_1}$$

in which g_m is the mutual conductance of the valve. This impedance, Z_i, which appears across the terminals of the oscillator, consists of a resistance R_i and a reactance X_i in parallel with the oscillator.

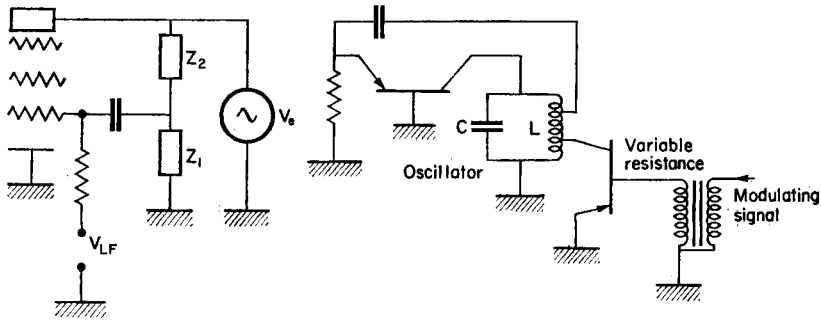


FIG. II.11. Reactance valve and reactance transistor.

Z_1	Z_2	R_i	X_i	
			inductive reactance for $X_i > 0$	capacitive reactance for $X_i < 0$
$\frac{1}{jC_1\omega}$	R_2	$\frac{1 + (R_2C_1\omega)^2}{g_m}$	$\frac{1 + (R_2C_1\omega)^2}{g_mR_2C_1\omega}$	
R_1	$\frac{1}{jC_2\omega}$	$\frac{1 + (R_1C_2\omega)^2}{g_m(R_1C_2\omega)^2}$		$-\frac{1 + (R_1C_2\omega)^2}{g_mR_1C_2\omega}$
$jL_1\omega$	R_2	$\frac{R_2^2 + (L_1\omega)^2}{g_m(L_1\omega)^2}$		$-\frac{R_2^2 + (L_1\omega)^2}{g_mR_2C_1\omega}$
R_1	$jL_2\omega$	$\frac{R_1^2 + (L_2\omega)^2}{g_mR_1^2}$	$\frac{R_1^2 + (L_2\omega)^2}{g_mR_1L_2\omega}$	

FIG. II.12. Effective impedances appearing across the oscillator.

Collected together in the table labelled Fig. II.12 are the various special cases, where Z_1 and Z_2 can each be either a capacitance or an inductance.

The working of this system depends upon the fact that the mutual conductance g_m, of the vacuum tube varies with the grid voltage and follows the applied LF. The maximum variation of frequency depends on the degree of linearity which is desired.

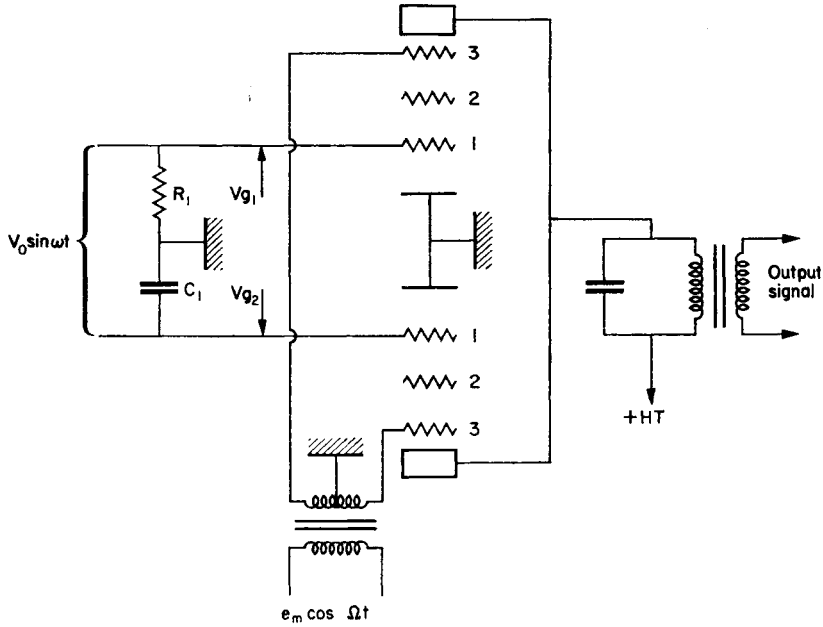


FIG. II.13. Phase modulator.

II.2.2.3. *Phase Modulation Devices*

Two methods of producing phase modulation are described – the Armstrong modulator and the serasoid modulator.

In the Armstrong method, two vectors initially in quadrature are combined after one or both vectors have been modulated in amplitude by the LF signal. A modulation device of this kind is shown in Fig. II.13. The maximum phase variation which can be obtained is given by

$$\varphi_{\max} = \tan^{-1}\left(\frac{\sqrt{2}e_m}{V_0}\right)$$

in which e_m is the amplitude of the LF signal which is applied to the grids and V_0 is the amplitude of the HF signal which is applied to the grids. This system has the drawback of creating a parasitic amplitude modulation, the percentage effect of which is given by

$$M = \frac{1 - \cos \Delta\varphi}{1 + \cos \Delta\varphi}.$$

To prevent this parasitic modulation from becoming troublesome, a small value of maximum phase variation is chosen; normally, we take $\Delta\varphi_{max} \leqslant 0{\cdot}2$ radians, so that the percentage amplitude modulation is very slight, and given by

$$M \simeq \frac{(\Delta\varphi_{max})^2}{4}.$$

Another way of obtaining phase modulation is provided by the serasoid arrangement. With this system, large phase deviations can be attained directly, of at most 150°. Nevertheless, it is usual to accept $\pm 90°$ as the limits of the phase variation, in order to avoid any parasitic amplitude modulation, and to obtain a satisfactory linearity. A highly stable quartz crystal oscillator is used as the local oscillator. Its output is shaped, by squaring, differentiating and clipping to produce pulses at the local oscillator frequency, and these are used to trigger a saw-tooth generator. The saw-tooth, which is highly linear, is then summed with the LF signal at the input to a valve or transistor stage. By suitable biasing it is arranged that in the absence of the LF modulating signal the valve limits at the mid-level of the saw-tooth amplitude.

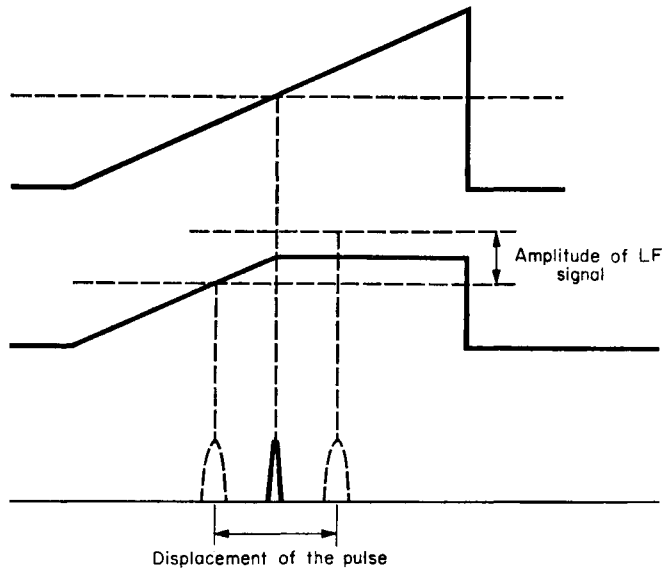


FIG. II.14. The principle of a serasoid device.

In the presence of the LF voltage the limiting level is thus decreased or increased, and by differentiating the limited saw-tooth a pulse results each cycle of which is advanced or retarded in time as shown in Fig. II.14.

Thus the amplitude and sign of the phase change with respect to the local oscillator is proportional to the level of the applied LF. A semi-diagrammatic form of the complete serasoid modulator is shown in Fig. II.15.

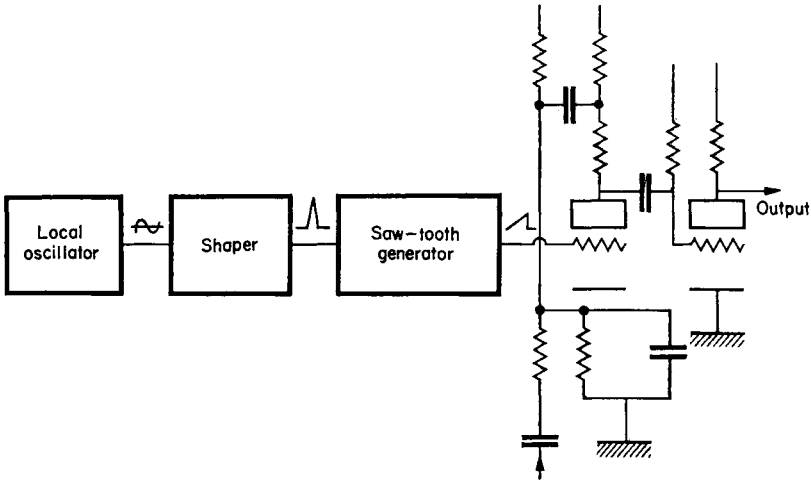


FIG. II.15. A serasoid device.

II.2.3. *Reception of FM and PM Waves*

In frequency modulation, amplifier linearity is of prime importance, as the phase distortion which could otherwise appear would produce perturbations of the FM signal by variations of the group period, $\tau = d\varphi/d\omega$. There are, however, two elements, peculiar to FM reception, which are nonlinear and which it is convenient to treat in a special way. These are the limiter and discriminator stages.

II.2.3.1. *Limiters*

These devices are designed to suppress the parasitic amplitude modulation, and to remove the peaks of the parasitic noise disturbances to improve the signal-to-noise ratio.

Whatever the level of the input signal, the output amplitude delivered by the limiter must be constant to allow the discriminator to function normally. A typical limiter characteristic is drawn in Fig. II.16. The output provided by this stage is constant only if the input amplitude exceeds a certain value termed the "limiter threshold".

The signal which is obtained at the output of the limiter is rich in harmonics, and a circuit tuned to the fundamental frequency eliminates the unwanted components.

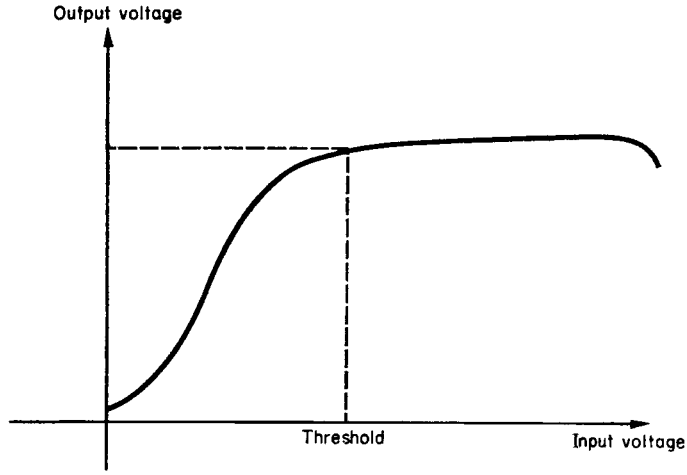


FIG. II.16. Typical characteristics of a limiter.

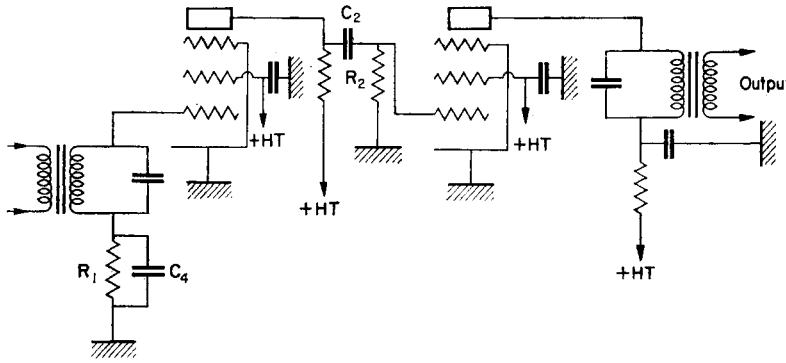


FIG. II.17. Double limiter circuit.

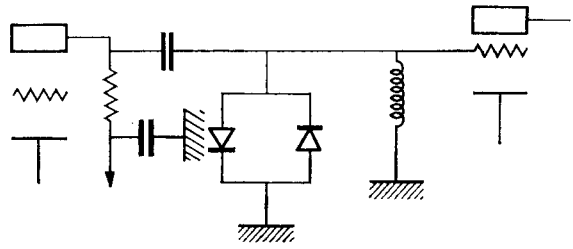


FIG. II.18. Double diode limiter.

Many types of limiters are in existence, such as grid current limiters and diode limiters, among others. The grid current limiter (see Fig. II.17) consists of a pentode with either the anode potential or the screen potential set at a low value. As soon as the excitation potential is sufficient for the grid-cathode voltage to reach zero, an autobias current is set up in the grid circuit. The choice of time constant, RC, of the grid circuit is dictated by the frequency of the parasitic disturbances which one is attempting to eliminate. If, however, the time constant is reduced, the height of the threshold increases and the level of signals for which the limiter is effective is also reduced. For this reason two limiters are very often used in cascade; the time constant of the first is shorter than that of its successor.

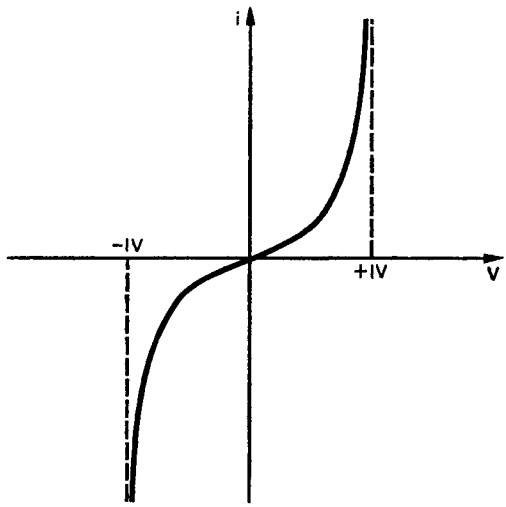


FIG. II.19. Current-voltage characteristic of a pair of diodes connected in parallel opposition.

Methods of limiting with diodes are of particular interest, since they save a valve and occupy little space. In a double-diode limiter the two valves are connected in parallel in opposite sense as shown in Fig. II.18. The diode current is an approximately exponential function of the applied voltage and Figure II.19 shows the variation of the current in the two diodes as a function of the applied voltage.

With diodes like the 19P2, the curvature is very marked, and the curve $i = f(V)$ can be regarded as tangential to the asymptotes $V = \pm 1$ volt. This means that the limiter threshold is one volt.

To ensure the diodes are driven rapidly into limiting conditions the driving valve should have a relatively high anode current availability and a high stage gain. The system has no appreciable time constant, since smoothing is not employed. The diodes themselves are frequency sensitive, and their choice is thus affected by the HF employed.

II.2.3.2. *Discriminators*

The role of a discriminator is to extract the modulating signal from the FM wave. Numerous devices which can perform this operation are available; only diode discriminators which are the type most frequently met with, are described here.

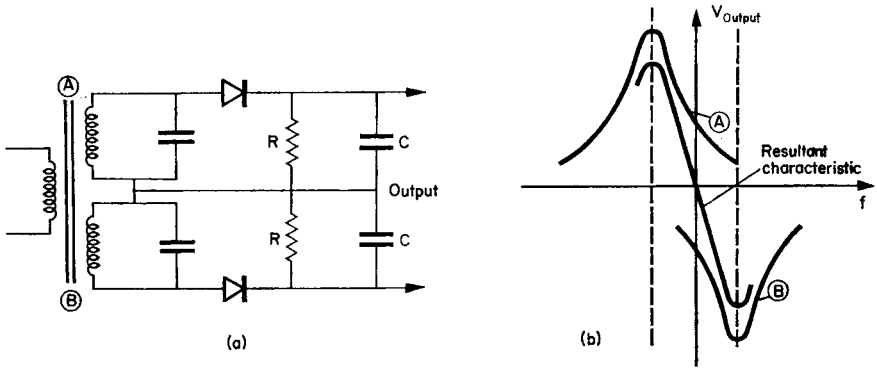


FIG. II.20. The Travis discriminator.

The "Travis" discriminator, shown in Fig. II.20, consists of two circuits tuned to frequencies displaced with respect to one another. In Fig. II.20b are plotted the selectivity curves (A) and (B) corresponding to circuits (A) and (B) of Fig. II.20a. The output voltage is the algebraic sum of the two

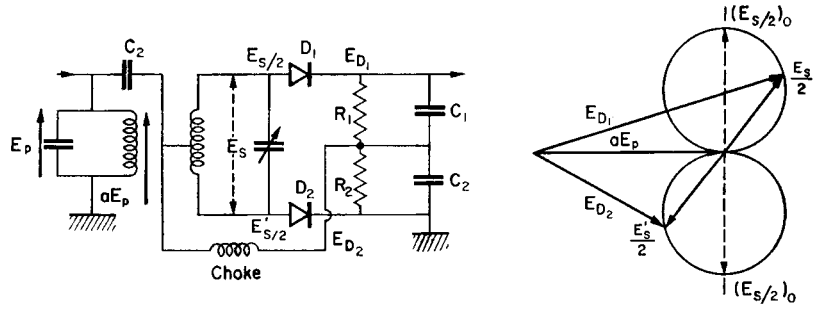


FIG. II.21. The Foster–Seeley discriminator.

circuit voltages, that is, the arithmetic difference between the voltages across (A) and (B). If the difference between the resonant frequencies of the two circuits is $1{\cdot}5f_0/Q$, the resulting characteristic is linear to better than 1 per cent.

Another type of discriminator which is very frequently employed is the Foster–Seeley version; the principle is shown schematically in Fig. II.21.

A voltage aE_P, which is in quadrature with the primary voltage E_P, is applied to the mid-point of the secondary of the transformer. The voltage of the latter, E_s, can split into two voltages of opposite phase, $E_{s/2}$ and $E'_{s/2}$. When the system is tuned, $E_{s/2} = (E_{s/2})_0$ and $E'_{s/2} = (E'_{s/2})_0$; the voltages E_{D1} and E_{D2} at the outputs of the two diodes are equal and the discriminator output voltage is zero. When the applied frequency differs from the tuned frequency, the two voltages E_{D1} and E_{D2} are no longer equal and, as the vector diagram in Fig. II.21 shows, a voltage appears at the output of the discriminator.

A typical example of a transistorized limiter–discriminator circuit is illustrated in Fig. II.22.

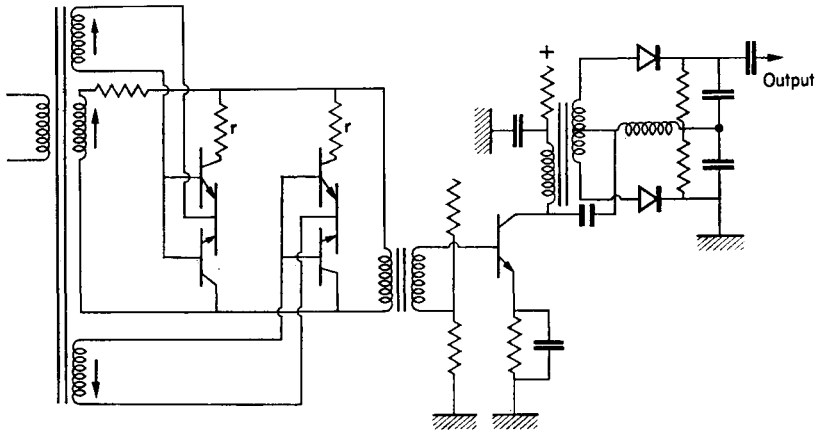


FIG. II.22. Transistor limiter–discriminator circuit[85].

II.3. PULSE MODULATION

Five kinds of pulse modulation can be distinguished: pulse amplitude modulation (PAM), pulse duration modulation (PDM), pulse position modulation (PPM), pulse frequency modulation (PFM), and finally pulse coded modulation (PCM).

The features of PAM, PDM, PPM and PFM modulation are first summarized together, and we conclude this Chapter with a more detailed study of the PCM system, whose applications to remote-control and telemetry are of special interest.

II.3.1. *The Spectrum of a Train of Pulses. The Case of Modulated Pulses*[10,53]

Consider the train of pulses of Fig. II.23; if $f_r = 1/T_r$ is the repetition frequency, and the width of each pulse is θ, then a duty factor δ can be defined as the ratio $\delta = \theta/T_r$.

The signal $f(t)$ of Fig. II.23 can be expressed analytically by a series of the form

$$f(t) = \delta A \left[1 + 2 \sum_{n=1}^{\infty} \frac{\sin n\delta\pi}{n\delta\pi} \cos n\Omega_r t \right]$$

in which Ω_r is the angular repetition frequency of the pulses. The spectrum of such a signal is clearly infinite in extent since it contains spectral components of frequency $0, f_r, 2f_r, \ldots, nf_r$. If the repetition frequency and the

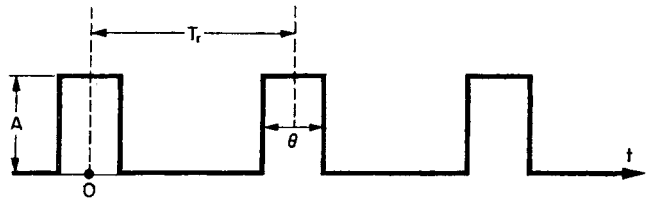


FIG. II.23. Pulse train.

shape of the pulses remain constant, the HF wave modulated by this signal is of the form

$$V(t) = \delta A \cos \omega t \left[1 + 2 \sum_{n=1}^{\infty} \frac{\sin n\delta\pi}{n\delta\pi} \cos \Omega_r t \right],$$

in which n is a positive integer.

When the pulse train is modulated in amplitude in such a way that the height of the pulses is proportional to the modulating signal, $s(t)$, we have

$$V_{AM}(t) = V(t) [1 + m \sin \Omega t],$$

in which m is the modulation factor and Ω is the angular frequency of the modulating signal.

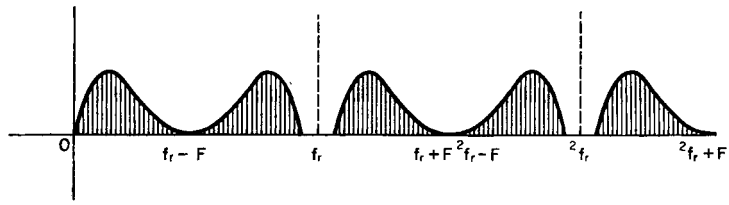


FIG. II.24. Spectral distribution of a PAM signal.

The spectrum of the PAM wave (see Fig. II.24) therefore consists of
– a constant component of amplitude δA
– a component with the amplitude modulation frequency, $m\delta A$

– components which correspond to multiples nf_r of the repetition frequency with amplitudes $2A\delta \dfrac{\sin n\delta\pi}{n\delta\pi}$

– and finally, components which correspond to the frequencies $nf_r \pm F$, with amplitudes $m\delta A \dfrac{\sin n\delta\pi}{n\delta\pi}$.

At the receiver, the modulating signal is recovered by means of a low-pass filter.

To enable the filtered portion of the spectrum to be isolated correctly, the partial spectra must not overlap one another, and this implies that the condition of $F_{max} \leqslant f_r - F_{max}$ must be satisfied, or what is equivalent,

$$f_r \geqslant 2F_{max}.$$

The repetition frequency of the pulses has thus to be at least twice the maximum LF frequency which is to be transmitted.

We shall now consider the case of a PPM wave; the pulses occur at times t_n such that

$$t_n = nT_r + \Delta T_r \sin \Omega t_n.$$

To prevent the pulses overlapping and to ensure, therefore, that there is only one in each period, requires

$$\Delta T_r < \frac{T_r}{2}$$

in which ΔT_r represents the maximum displacement of a pulse which the modulating signal creates. It can be shown that the complex PPM signal consists of

– a constant component, the amplitude of which is $A\delta$

– a component with the signal frequency F and amplitude

$$A\Omega\delta\,\Delta T_r \frac{\sin \dfrac{\Omega\theta}{2}}{\dfrac{\Omega\theta}{2}}$$

which is in practice equal to $A\Omega\,\delta\Delta T_r$, since $\delta \ll 1$ usually

– a series of components with frequencies nf_r, and amplitudes

$$2A\delta \frac{\sin \dfrac{n\Omega_r\theta}{2}}{\dfrac{n\Omega_r\theta}{2}} J_0(n\Omega_r \Delta T_r)$$

– a series of components with frequencies $nf_r \pm kF$ and amplitudes

$$2A\delta \frac{\sin\left[\frac{n\Omega_r \pm k\Omega}{2}\right]}{n\Omega_r} J_k(n\Omega_r \Delta T_r).$$

The component with the signal frequency, $A\Omega\delta\Delta T_r$, usually has a very small amplitude ($\delta \ll 1$ and ΔT_r is small), and this makes recovery of the LF signal at the receiver more complicated. It is preferable to transform the PPM pulses into PDM pulses.

In the case of PDM pulses, it is the quantity θ which is modulated,

$$\theta = \theta_0(1 + m \sin \Omega t).$$

If the mean duration of the pulses in zero, if, that is, it is arranged that the positions of the unmodulated pulses coincide with the mean positions of the modulated pulses, we have $\theta = m\theta_0 \sin \Omega t$.

The spectrum of a PDM signal consists of

– a constant component

$$A\delta_0$$

– a component at the signal frequency

$$mA\delta_0$$

– components at the frequencies nf_r

$$\frac{2A}{T_r}[1 + J_0(n\Omega_r m\theta_0)]$$

– components at the frequencies $nf_r \pm kF$

$$2Am\delta_0 \frac{J_k(n\Omega_r m\theta_0)}{n\Omega_r m\theta_0}.$$

In the case of a PDM or a PPM signal, each component of frequency nf_r is surrounded by a series of equidistant lines, at frequencies

$$nf_r \pm F$$
$$nf_r \pm 2F$$
$$\vdots$$
$$nf_r \pm kF$$

In the same way, the constant component is surrounded by the LF components $F, 2F, \ldots, kF, \ldots$

If the repetition frequency is much higher than the maximum LF frequency which is to be transmitted, the partial spectrum which contains all the information in the signal can be isolated by means of a low-pass filter.

II.3.2. *Bandwidth*

The energy of the rectangular wave represented by

$$f(t) = \delta A\left[1 + 2\sum_{n=1}^{\infty} \frac{\sin n\delta\pi}{n\delta\pi} \cos n\Omega_r t\right]$$

can be divided as follows:
– the mean power which is proportional to δA^2
– the power carried by the component of zero frequency which is $\delta^2 A^2$
– the power carried by successive harmonics which is proportional to

$$4\delta^2 A^2 \sum_{n=1}^{\infty} \left(\frac{\sin n\delta\pi}{n\delta\pi}\right)^2$$

or

$$\delta^2 A^2 \sum_{n=-\infty}^{\infty} \left(\frac{\sin n\delta\pi}{n\delta\pi}\right)^2, \quad \text{excluding } n = 0.$$

In theory, $1 < n < +\infty$, and the spectrum extends to infinity, and thus the bandwidth occupied by the signal is infinite. The practical bandwidth, however, we shall regard as the width of the band which contains 96·6 per cent of the energy.

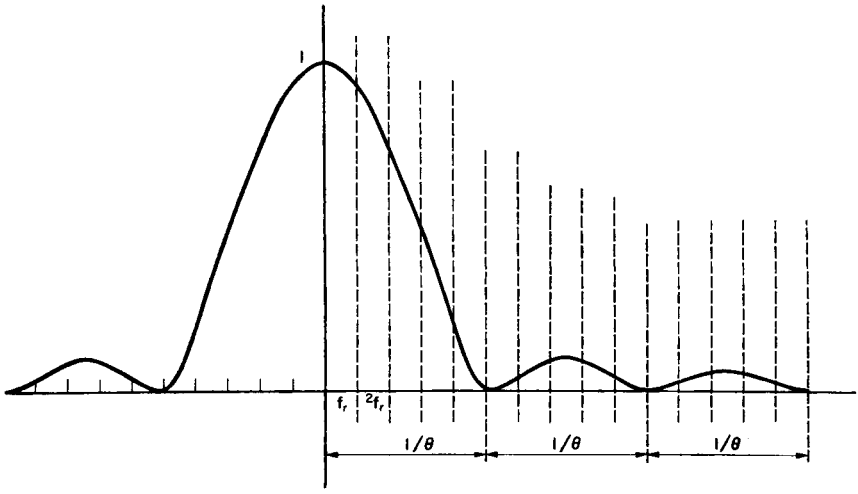


FIG. II.25. The power in successive harmonics of the repetition frequency.

In Fig. II.25 is plotted the curve $\left(\frac{\sin n\delta\pi}{n\delta\pi}\right)^2$ as a function of the frequency f.

When there is modulation, the period T_r is no longer constant, and the envelope of Fig. II.25 is shifted to one side or the other of the mean position.

This is apparent as a blurring of the maxima and minima, but does not complicate the situation further since the spectrum of the rectangular wave, whether it is modulated or not, can be regarded as remaining unchanged. The first peak contains 90 per cent of the energy, the second 5 per cent and third 1·66 per cent. Thus a frequency band of $\pm 3/\theta$ contains 96·6 per cent of the total energy.

HF or MF bandwidths at the receiver of $2/\theta$, $4/\theta$ or $6/\theta$ can be admitted according to the distortion factor which can be tolerated. The shape of the pulse selected is seen to govern the useful bandwidth; the narrower the pulse (small θ) the wider becomes the band.

The optimum pulse-shape for which the minimum band-width is obtained, is a Gaussian curve.

When the repetition frequency is not markedly greater than the maximum frequency to be transmitted, the value of the bandwidth can be taken as

$$\mathscr{B} = 6\left(\frac{1}{\theta} + F_{\max}\right) = 6\left[\frac{f_r}{\delta} + F_{\max}\right].$$

II.3.3. *Modulation of the Pulses*

The pulses are usually produced by means of a multivibrator. The leading edges of the square wave are differentiated to obtain such pulses. An example of a transistorized pulse generator using a blocking oscillator is given in Fig. II.26.

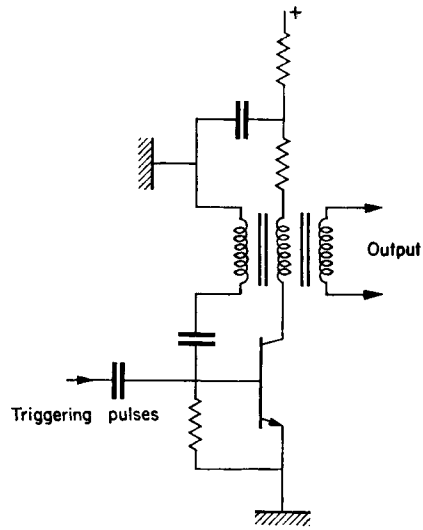


FIG. II.26. Blocking oscillator.

II.3.3.1. *Pulse Amplitude Modulation (PAM)*

The diode modulator illustrated in Fig. II.27 is frequently employed. The pulse is modulated symmetrically by the LF signal. Figure II.28 shows another modulation system in which a pentode valve is used as modulating

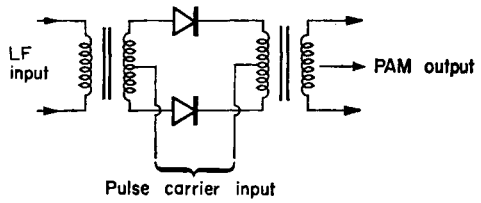


FIG. II.27. Diode modulator.

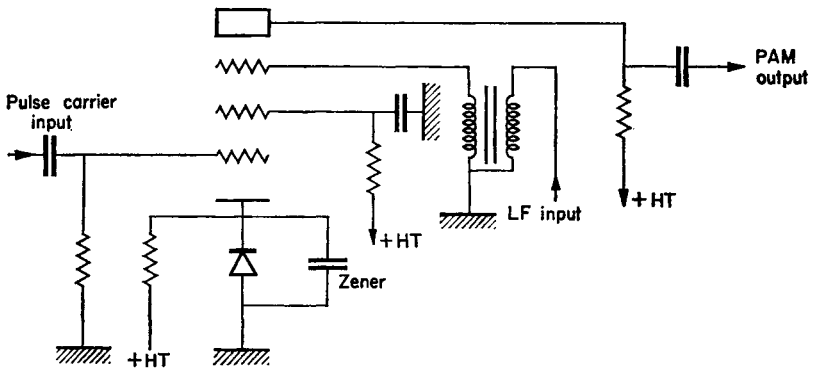


FIG. II.28. Valve modulator.

tube. The suppressor grid operates as a second control grid, to which the modulation signal is applied. The bias of the cathode must be adequate to block the tube in the absence of any applied pulse; this can be achieved by inserting a Zener diode with an adequate reference potential into the cathode circuit.

II.3.3.2. *Modulation of the Pulse Duration (PDM)*

A PDM wave can be obtained by making use of the serasoid principle discussed earlier. The unmodulated saw-tooth is shown in Fig. II.29, and the presence of the LF moves the pulse *A* along the time axis in synchronism. If pulse *A* is now used to set a bistable multivibrator and pulse *B*, which is fixed in time, to reset it, we obtain at the output of the device a rectangular

wave whose duration is proportional to the interval AB and hence to the modulating amplitude.

A different device, illustrated in Fig. II.30, converts an amplitude-modulated pulse into a pulse modulated in duration.

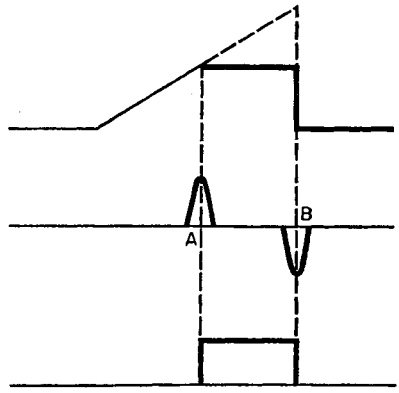


FIG. II.29. Principle of PDM by the serasoid method.

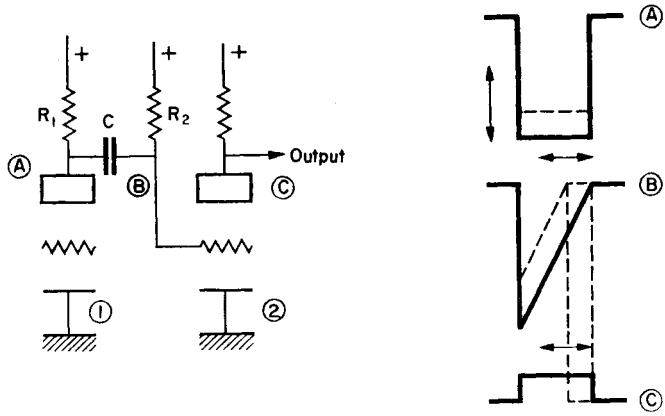


FIG. II.30. PAM–PDM converter.

When no signal is being applied to valve 1, the grid of valve 2 is positively biased with respect to the cathode, and this second valve behaves like a diode. When a pulse appears at A, the second valve is cut off, and the capacitor C discharges into the circuit $(R_1 + R_2)$. The time taken by the capacitor to discharge depends upon the height of the applied pulse, and thus a PDM wave appears at the anode of valve 2.

II.3.3.3. *Pulse Position Modulation (PPM)*

A PPM wave can again be obtained directly with the aid of the serasoid arrangement. In this case only the pulse A of Fig. II.29 is used, modulated in position in synchronism with the LF, as has already been seen.

It is equally feasible to employ a device which converts PDM into PPM. It is only necessary to apply a PDM wave to the input of a differentiator, which results in two pulses corresponding to the front and rear vertical edges of the PDM pulse. By retaining only one of these two pulses, a PPM wave is obtained. The polarity of the pulse which corresponds to the fixed face is the opposite of that of the pulse which is produced by the variable falling edge, and can thus be easily eliminated.

II.3.3.4. *Detection of the Modulated Pulses*

In the discussion of spectra it was seen that there is always a component corresponding to modulating frequency. Accordingly a low-pass filter can be placed at the output of the detector, arranged to reject the repetition frequency of the pulses and its harmonics, but to accept the maximum modulating frequency without distortion.

II.4. CODED MODULATION (PCM)

The signal, properly speaking, is now *quantized* as we shall see in Chapter III; by this is meant that by sampling the signal, some chosen level in a discontinuous series is substituted for the signal in such a way that the error between the level selected and the real value does not exceed the error in the quantization.

It is possible to translate any number in a binary form containing only two possible digits, "0" and "1". The essential advantage of yes-no transmission (presence or absence of a pulse) lies in the efficiency of the protection it offers against parasitics. The signal-to-noise characteristics of this type of modulation will be studied in the chapter entitled "Limitations on Information".

II.4.1. *Evaluation of the Bandwidth in PCM*

If F_m is the maximum frequency of the modulation which is to be transmitted, then the sampling frequency must be greater than or equal to $2F_m$ (see II.3.1).

In a code of n binary digits the maximum number of pulses corresponding to the maximum number of digits is n; allowance must also be made for the number of synchronizing pulses. The minimum repetition frequency of the pulses then becomes $2(n + s)F_m$.

When the code pulses are rectangular, the bandwidth extends in theory to infinity, and practical bandwidths are excessively large.

To benefit fully from the advantages of PCM modulation, pulses which have particularly easily reproducible shape, are employed. Theory suggests and experiment confirms that when a function is of zero slope at the origin, and is elsewhere rounded off, the spectrum obeys a law in $1/f^n$ $(n \geqq 3)$.

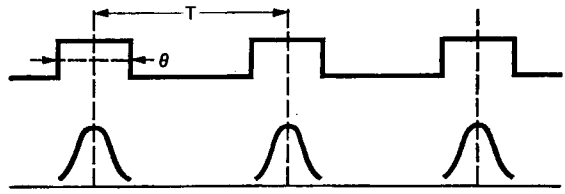


FIG. II.31. Train of code pulses.

When the function is gaussian $e^{-\pi t^2}$, the spectrum is of the form $e^{-\pi f^2}$. For the pulse-train shown in Fig. II.31, the analytic expression for the gaussian pulse is then

$$s(t) = e^{-\alpha t^2/T^2}.$$

It is arranged that $s(\theta/2) = 1/2$, which defines the value of α. The spectrum of this Gaussian pulse is equal to

$$S(\omega) = e^{-\omega^2 T^2/4\sqrt{\alpha}}.$$

If the spectrum is truncated at a value corresponding to the cut-off angular frequency ω_c, for which $S(\omega_c) = e^{-2}$, we obtain

$$f_c = \frac{\sqrt{2}\,(\alpha)^{1/4}}{\pi} \times \frac{1}{T},$$

$$f_c \simeq \frac{0{\cdot}6}{T}, \quad \text{with} \quad T = \frac{1}{2(n + s)F_m}.$$

As a result, the minimum bandwidth which is capable of accepting the above pulses is equal to

$$\mathscr{B} = f_c = 1{\cdot}2(n + s)F_m.$$

The pulses can be allowed to overlap without fear of cross-modulation since in coded modulation, the data within the signal are represented by the presence or absence of pulses. This point deserves to be emphasized; it is one of the essential reasons for the advantage of coded modulation over the other types of modulation. In a more general way, we may say that the pulse train is represented by a square wave, the frequency of which is equal to half the information rate.

II.4.1.1. *Bandwidth for a PCM–FM Transmission*

It is assumed that the frequency of the wave remains unaltered when a digit "0" is applied to the HF wave of frequency f, but that the appearance of a digit "1" does affect the frequency. A square wave can be expanded as a Fourier series of the form:

$$S(x) = \frac{4}{\pi}\left[\cos x - \frac{1}{3}\cos 3x + \frac{1}{5}\cos 5x - \frac{1}{7}\cos 7x + \cdots\right].$$

It is clear, therefore, that the bandwidth required for a pure square wave is too wide. However there is no loss of information if it is filtered and only the fundamental, FM wave, retained. The response characteristic of a suitable type of filter is shown in Fig. II.32, in which the cut-off frequency f equals half the transmission rate.

When this filter is employed, the pulse train satisfies the conditions necessary to occupy the minimum bandwidth. The HF bandwidth is now easy to calculate. If R is the transmission rate in bits/second, the maximum frequency to be transmitted is $F_m = R/2$ c/s. For a modulation coefficient m, we have

$$\mathscr{B} = 2(m + 1)F_m.$$

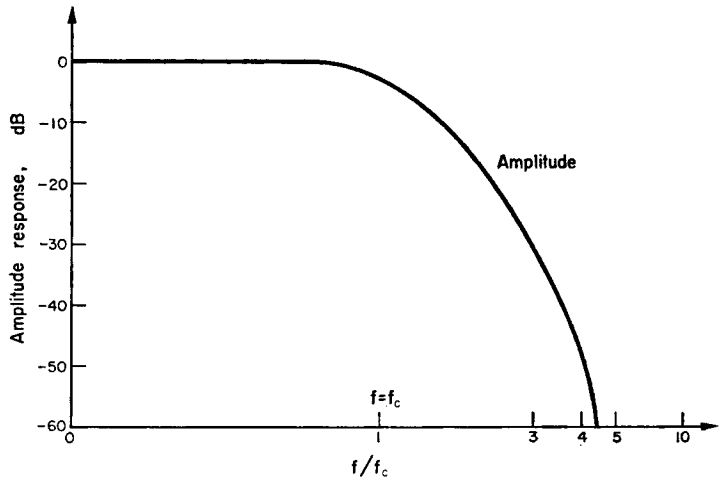


FIG. II.32. Response curve of the filter (27).

It is interesting, too, to know the 60 db bandwidth. On the curve of Fig. II.32, the attenuation is seen to be 60 db when $F = 4{\cdot}7\,F_c$. The bandwidth at 60 db, therefore, is given by

$$\mathscr{B}_{60db} = 2(\Delta f + 4{\cdot}7 f_c).$$

Figure II.33 shows the value of the bandwidth at 60 db as function of the information rate for different values of the frequency change.

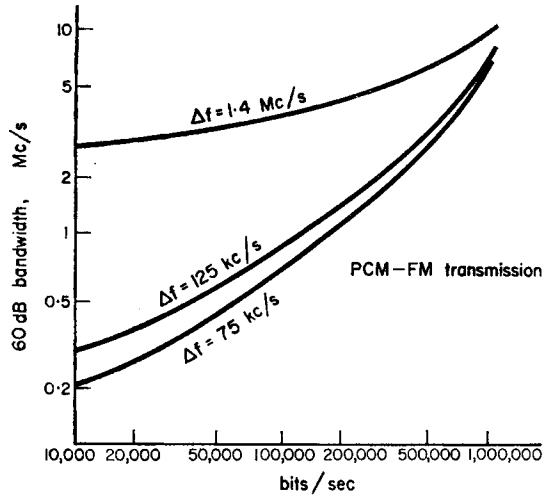


FIG. II.33. HF 60 db bandwidth as a function of the information rate.

II.4.1.2. *The Bandwidth for PCM–AM Transmission*

The choice of the parameters can be made more simply in this case than for PCM–FM modulation, since the HF band only comprises the two ordinary sidebands of amplitude modulation. The HF bandwidth therefore, takes the value

$$\mathscr{B} = 2F_m.$$

II.4.2. *Synchronization*

When several channels are commutated, the synchronization pulse serves as a reference, so that pulses can be separated at the receiver. We shall return to this problem in the section on multiplex systems.

In the case with which we are more particularly concerned, the synchronization separates the code groups, provides a reference for the time zero, and indicates where words begin or where the code group corresponding to the word starts. Finally, synchronization is indispensable at the receiver, to distinguish between the presence and the absence of pulses, to distinguish a digit "0" from a digit "1".

The synchronization signal can be transmitted through a special channel, or alternatively, detected by means of the actual coding arrangement itself, or again, marked out by a special modulation.

One solution which may be adopted is to begin a code group with a digit "1" and to end with a "0". In the decoding chain an element is inserted which extracts the presumed synchronization reference and a decision circuit checks whether the reference is acceptable as satisfactory. In the least favourable case, for n digits per code group, the decision is taken at the $(n - 1)$-th.

The simplest precedure consists in choosing a code pulse which is modulated yes-no at a frequency half that of the sampling frequency. The synchronization will then be performed at the receiver by means of a low-pass filter.

II.4.3. *Transmission*

An analogue signal can be transformed into a digital signal, as described in Chapter III on "Information and Coding".

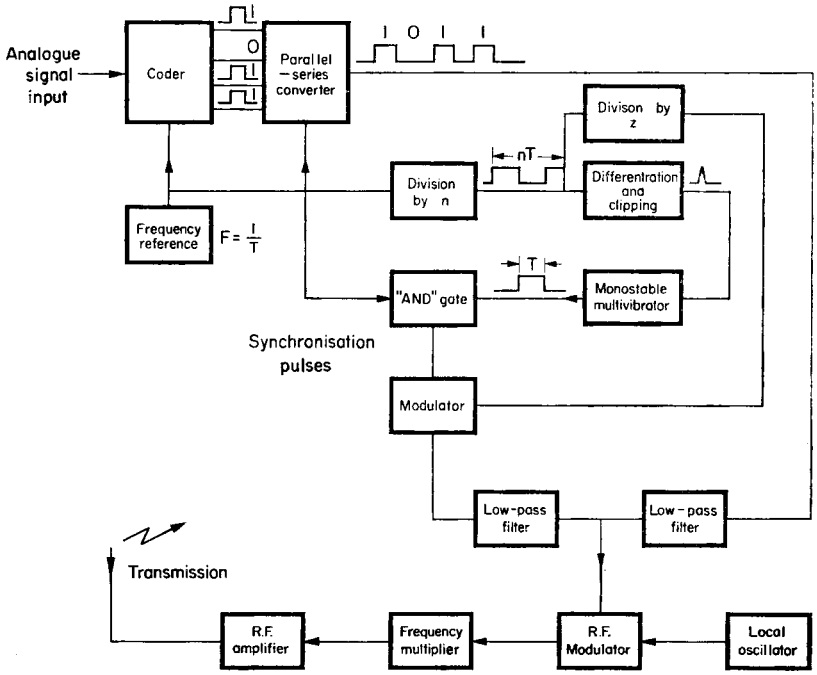


FIG. II.34. Block diagram of complete PCM–AM or FM transmitter.

Accordingly, analogue signals can be applied to the coder which turns the information in the signal into a digital form, so that at the output of this element the information appears translated into a parallel code. A parallel–series code converter gives the information in the form of a succession of pulses which are applied to a low-pass filter to eliminate the harmonics,

allowing a reduced bandwidth. The synchronization signal is formed by dividing down the local oscillator or clock frequency. The synchronization signal is then modulated in a manner which can be detected easily at the receiver. The resulting total signal is then used to modulate an RF frequency in AM or FM according to the type of modulation which has been selected. Figure II.34 is a diagram of such a transmitter.

The low-pass filter circuits strongly influence the behaviour of the signal.

Consider the series of digits 00110110, shown schematically in Fig. II.35. It will be seen that the distortion of the signal depends on the bandwidth allocated to it.

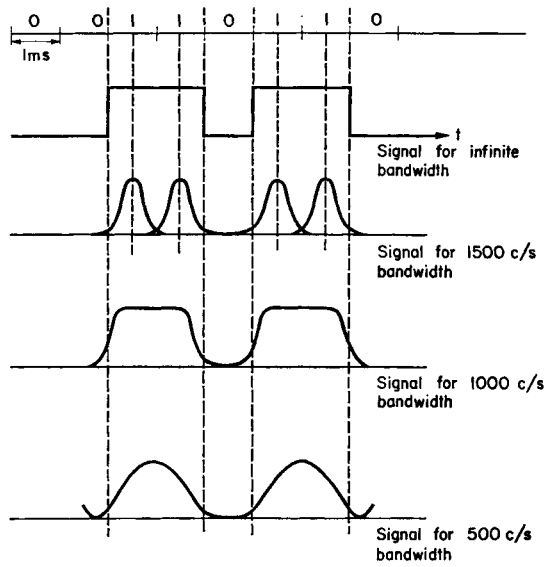


FIG. II.35. Distortion of signal passed through a low-pass filter.

The passage of a digit "1" is represented by a rising edge, whereas the passage from a digit "1" to a digit "0" is characterized by a falling edge. The wider the bandwidth allocated, the more sharply is the edge determined, but for reasons of economy, one tends to choose narrow bandwidths. Another interesting point is that with a bandwidth equal to $1{\cdot}5f_0$, the value of the corresponding digit can be detected simply by coincidence between the amplitudes. It should also be noted that when a pulse passes suddenly through a filter, a phase change which may be equal to $\pi/2$ may occur. It is thus necessary to watch out for these phenomena – at the receiver in particulary – in order to be able to make any necessary correction for this phase change. Above all, it is important to ensure that the synchronization pulse itself also undergoes this same phase change, as otherwise there is a risk of errors occurring at the decoding stage.

II.4.4. *The Receiver*

The receiver system contains first of all an ordinary RF receiver unit, with its local oscillator, its mixer, and its intermediate frequency. The demodulation unit consists of a discriminator, in the case of frequency modulation,

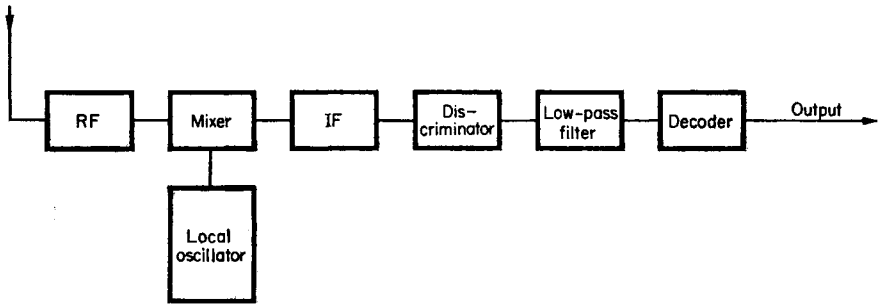


FIG. II.36. Block diagram of a receiver.

or of an ordinary detector for amplitude modulation, followed by a low-pass filter to eliminate unwanted components. Finally, there is the decoding unit for this type of PCM modulation (see Fig. II.36).

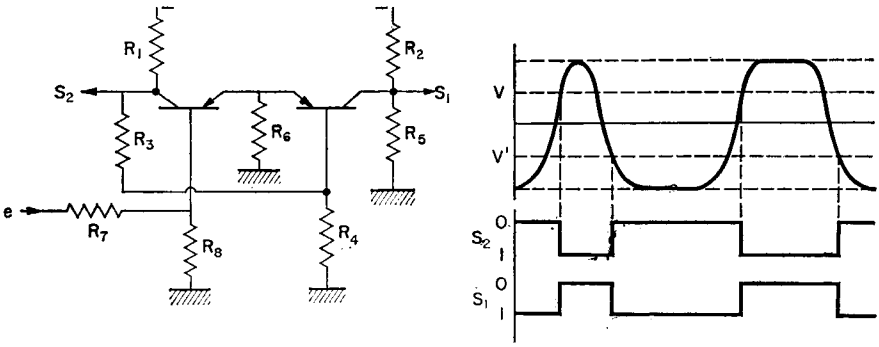


FIG. II.37. Schmitt trigger.

The decoding unit is now considered in more detail. It consists of a synchronization detector, coincidence circuits for reading out the code, and a digital translator—the analogue of the one which provides the transmitted signal. Code pulses which are detected are then shaped by regenerative circuits (see Fig. II.37) or triggers.

This circuit depends on the coupling between the collector of $T1$ and the base of $T2$. The trigger changes its state as soon as the input signal goes out-

side a critical zone, $(U' - U)$. It is preferable to set the interval $(U' - U)$ to a small value, by means of the potential divider R_7R_8. In practice, the critical zone is arranged to lie equidistant from the maxima and minima of the signal.

It is now necessary to detect the synchronization pulse, knowing it to be modulated at a frequency which is half the repetition frequency. A pulse with frequency $f_0/2$ is detected and then doubled. Finally, a square wave results, whose leading edge is used to trigger a multivibrator.

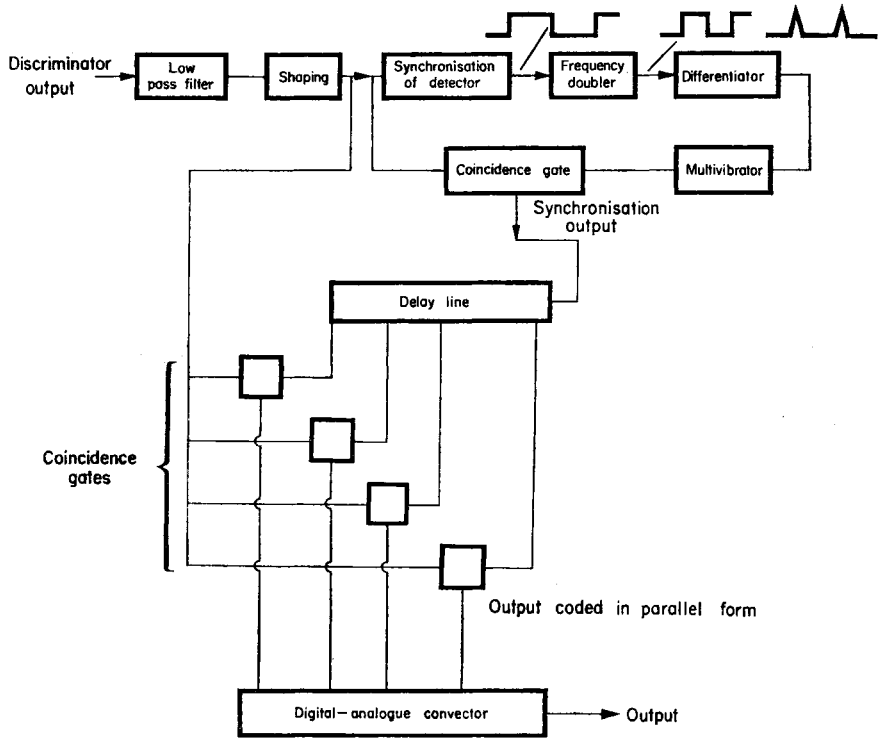


FIG. II.38. Block diagram of a decoder.

After bringing this wave into coincidence with the pulse train, a synchronization pulse is obtained which is pure enough to be used to "strobe" time instants in the code as is seen in Fig. II.38. A digital converter provides the required signal at its output.

The coincidence circuits perform the logical operation of combination "AND"; in other words, if X and Y are two pulses applied simultaneously to the "AND" circuit, the output delivers a pulse only if the product XY is equal to 1, which occurs when the X and Y are both the digits "1".

Such a circuit, employing diodes, is shown in Fig. II.39.

If the pulses X and Y are equal to E, no current flows in the resistor R, and the voltage at both ends of R is E. If either X or Y is zero, a current flows through the resistor, and creates a voltage drop which is sufficient to reduce the output voltage to a low level.

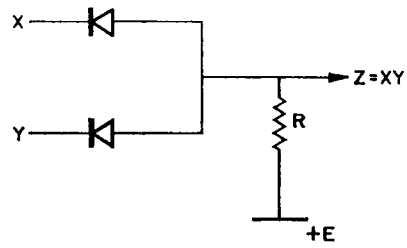


FIG. II.39. A diode "AND" logic circuit.

APPENDIX — GAUSSIAN PULSES[100]

In order to obtain a high signal-to-noise ratio, while using only a relatively narrow bandwidth, one is led to use pulses of a special shape. It will be shown that the signal-to-noise ratio takes its optimum value when the form of the signal which is received is the Fourier transform of the pass band, and this leads to the use of a Gaussian pulse.

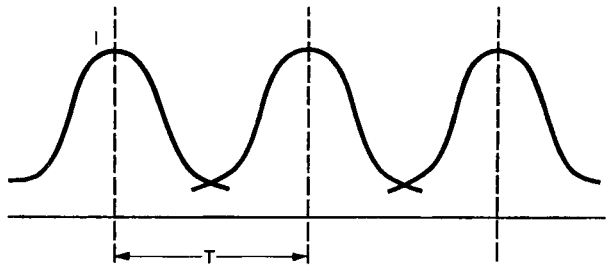


FIG. II.40. Gaussian pulses.

The expression for a Gaussian curve is

$$y = e^{-x^2}.$$

This curve is its own Fourier transform:

$$e^{-\pi t^2/\tau^2} = \tau \int_{-\infty}^{\infty} e^{-\pi\tau^2 f^2} e^{-2j\pi tf}\, df.$$

The spectrum of a pulse

$$u(t) = e^{-c^2/\tau^2}$$

is of the form

$$\varphi(f) = e^{-\pi^2\tau^2 f^2}.$$

Suppose that T is the time-interval between two pulses (Fig. II.40); the cross-modulation level is

$$D = 2\,e^{-(t/\tau)^2}$$

so that

$$\frac{T^2}{\tau^2} = \log_e 2D.$$

Further the frequency f_0 for which the amplitude of the spectrum is $1/k$ times its maximum value is such that

$$1/k = e^{-\pi^2\tau^2 f_0^2}$$

$$\pi^2\tau^2 f_0^2 = \log_e k.$$

The bandwidth, f_0, for which the pulse overlap is D can be obtained from the following expression:

$$T^2 f_0^2 = \frac{1}{\pi^2} \log_e 2D \log_e k.$$

Introducing the repeat frequency, $f_r = 1/T$, we find:

$$f_0 = \frac{f_r}{\pi} (\log_e 2D \log_e k)^{1/2}.$$

Thus for $D = -40$ db with respect to the minimum level 1, the bandwidth at 40 db is equal to $1{\cdot}57 f_r$.

CHAPTER III

INFORMATION AND CODING

THE DEVELOPMENT of communication systems is dependent upon the two following procedures: the search for components and basic circuits suitable for the equipment, and the analysis of the behaviour of these elements with a view to obtaining the optimum efficiency.

Information theory provides the link between mathematical considerations and the design of the actual transmitting equipment. We are confronted with three essential requirements:

- to obtain transmission systems which possess the maximum efficiency
- to show which method will provide the best compromise between maximum efficiency and any self-imposed limits
- to study the extent to which the effects of noise can be reduced.

III.1. THE NATURE OF THE INFORMATION

The most general kind of communication system consists of the following elements:

- a source of information which chooses the message which is to be transmitted
- a coder, which transforms the initial message into signals of a special kind, better suited to the transmission
- a transmitter
- a transmission line or communication channel, which constitutes the intermediate physical medium between transmitter and receiver
- a receiver
- a decoder which reconstructs the initial message
- the user, or other person for whom the message is destined, who receives the information contained in the message
- finally the presence of the noise which is the cause of all distortions of the signal must be mentioned.

A block diagram of a complete transmission system is shown in Fig. III.1. Communication is satisfactory if the message emitted by the information source is intelligible enough to be understood when it reaches the user, despite the distortion and noise which may have distorted the contents of the signal.

A measure of the information can be defined by the number of degrees of freedom available in choosing one message from among the others; the totality of possible messages constitutes message-space. In practice, this space is always limited (the temperature, for example, is limited to within a certain range).

As our unit of information, we shall take the alternative between two possible choices; thus in the binary system, we can choose between two symbols only: 0 and 1. This unit of information is known as a "bit".

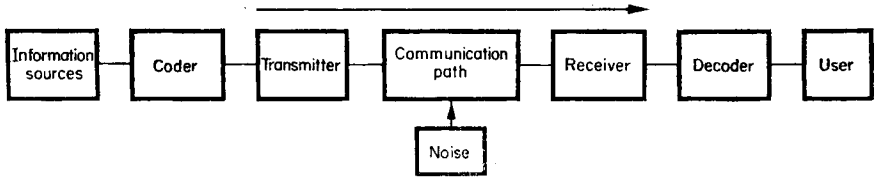


FIG. III.1. Block diagram of a complete transmission system.

In other words, the quantity of information measures the variation of possible events between two statistical states—initial and final—which result from the choice made by, for example, some instrument.

III.2. A MEASURE OF THE INFORMATION

As an example, consider the remote-control of a missile with four possibilities; by this, is meant that the missile will obey any of four orders: dive, climb, left and right. In this case the source has four symbols at its disposal.

The measure Q of the information in a source with M symbols at its disposal is

$$Q = \log M.$$

With the unit of information chosen above, the base of the logarithms is taken equal to 2

$$Q = \log_2 M.$$

If the communication consists of N symbols from among the M available the quantity of information which is transmitted is

$$Q = \log_2 M^N$$

or

$$Q = N \log_2 M.$$

The information density, or quantity or information per symbol is, therefore

$$H = \frac{Q}{N} = \log_2 M \quad \text{bits/symbol.}$$

More generally, if P_i is the probability belonging to choice of the i-th symbol, the mean information per symbol, H, or "entropy" is given by

$$H = -\sum_i P_i \log_2 P_i.$$

The entropy corresponding to an emitted symbol depends upon the probability distribution, P_i. It reaches its maximum when all the symbols are equally likely, provided they are wholly uncorrelated, that is, when

$$P_1 = P_2 = \cdots = P_i = \cdots \frac{1}{M},$$

so that

$$H_{\max} = -\sum \frac{1}{M} \log_2 \left(\frac{1}{M}\right)^M$$

or

$$H_{\max} = \log_2 M.$$

When the basic symbols are not all equally likely, and when correlations exist between them, the relative entropy is defined by the ratio $H/H_{\max} < 1$, and the redundancy as the difference between the relative entropy and unity

$$\text{Redundancy} = 1 - \frac{H}{H_{\max}}.$$

This quantity is a measure of the excess of information.

When the communication is noisy, when, that is, it is perturbed by the presence of errors and noise, the entropy conveyed through the system is less than that of the source of the messages.

In particular, when the information is transmitted in the form of voltages applied to circuits, the latter may limit the frequency band, and the superimposed noise in the passband is a cause of error.

The entropy of the signals received is equal to the difference between the entropy of the initial signals and that of the noise.

$$H_{\text{received}} = H_{\text{signal}} - H_{\text{noise}}.$$

The information rate R is the quantity of information transmitted per unit time,

$$R = \frac{Q}{t} = \frac{N}{t} \log_2 M \quad \text{bits/sec.}$$

If the N chosen symbols are emitted at a rate of n/second ($N = nt$), we obtain

$$R = n \log_2 M \quad \text{bits/sec.}$$

The information rate is a quantity which characterizes the possibility of transmitting a message in a given time.

III.3. THE CAPACITY OF A TRANSMISSION CHANNEL

The information in a message at the receiver is naturally limited. The spectrum of the signal in fact extends across a certain frequency band. The noise which arises in the transmission channel defines a threshold beneath which the data contained in the signal are no longer perceptible. In view of these limitations, we shall now examine the maximum information rate which can be attained in this channel. For this purpose, let us quantize the signal, and take into account the response time at the receiver.

First of all, it is necessary to know the number of different levels which the receiver can distinguish without error. In the absence of noise, the useful signal $f(t)$, varies between two amplitude levels, $+V$ and $-V$. Assume that onto this signal, noise is superimposed (see Fig. III.2).

The maximum amplitude of the noise fluctuations is $V = 2V_N$.

The instantaneous value of the signal is thus indeterminate to within $\pm V_N$. As a result, two instantaneous values of the signal must be separated by at least $2V_N$ if they are to be distinguishable.

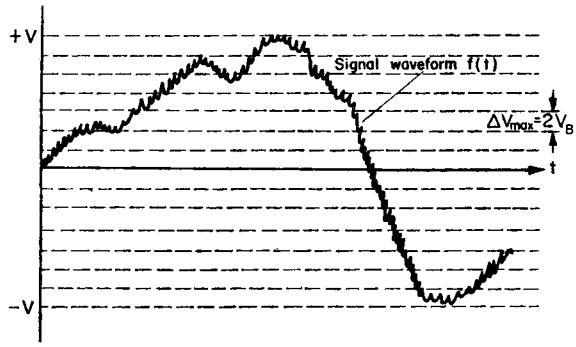


FIG. III.2. Signal with superimposed noise.

The number of perfectly distinct levels between the amplitudes $= +V$ and $-V$ is, therefore,

$$M = \frac{2V}{2V_N} = \frac{V}{V_N}$$

or, considering the powers of the signal P and of the noise N

$$M = \sqrt{\frac{P}{N}}.$$

We have carried out a quantization of the signal. The presence of the function $f(t)$ is therefore linked physically to the choice of ways of doing this,

given the M symbols which constitute the M distinct levels already obtained.

To be able to define the transmission potential of the system completely, it only remains to make allowance for the response time of the receiver.

In the most usual case, the receiver can always be represented by a band-pass filter, or alternatively, by changing the frequency, to a low-pass filter with a cut-off frequency f_c.

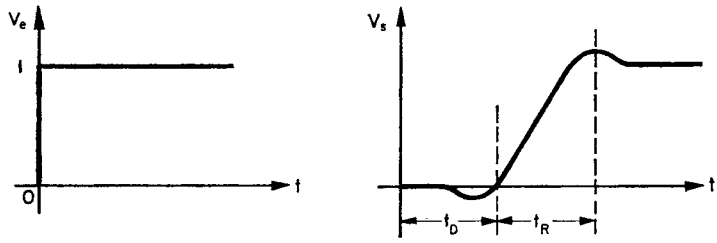


FIG. III.3. Response of a low-pass filter to a unit step.

Let us apply a signal V_e corresponding to a unit step to such a filter (see Fig. III.3). The wave V_s at the output of the low-pass filter has been delayed a time t_D with respect to the unit step, and only reaches its final value after a rise-time, t_M.

The shorter the filter characteristic, the longer becomes the delay t_D. The rise-time t_M, however, is related directly to the bandwidth of the filter, through its cut-off frequency f_c ($t_M = 1/2f_c$).

The result of this is that the time interval which corresponds to the passage from one level to another has to be greater than the rise time if the receiver is to be able to distinguish a sequence of successive levels, analogous to an emitted sequence of unit functions. In one second, therefore, a maximum of $n = 2f_c$ different signals can be distinguished, the latter being regularly spaced at $1/2f_c$ second intervals.

The maximum information rate is therefore

$$R_{\max} = n \log_2 M$$

or

$$R_{\max} = 2f_c \log_2 \sqrt{\frac{P}{N}} = f_c \log_2 \frac{P}{N} \quad \text{bits/sec.}$$

The power P is equal to the sum of the powers of the signal alone, S and of the noise, B.

$$R_{\max} = f_c \log \left(1 + \frac{S}{N}\right).$$

The preceding results can be expressed in the form of a well-known theorem, given by Shannon:

"The information rate R of a steady signal of mean power S, the spectrum of which occupies a bandwidth $\mathscr{B}$ and which undergoes the influence of noise the power of which is N, is less than or equal to a maximum value R_{max},

$$R \leqslant R_{max}$$

$$R_{max} = \mathscr{B} \log_2 \left(1 + \frac{S}{N}\right).$$

It should be noticed that the value R_{max} can only be attained if the spectra of the signal and the noise are of the same nature.

The capacity C of a transmission channel is the name given to the maximum information rate at which transmission can take place along the channel:

$$C = R_{max} = \mathscr{B} \log_2 \left(1 + \frac{S}{N}\right).$$

The quality of a channel is measured by its information capacity. In Fig. III.4 the expression $C/\mathscr{B}$ in bits/cycle is plotted as a function of the signal-to-noise power ratio in db.

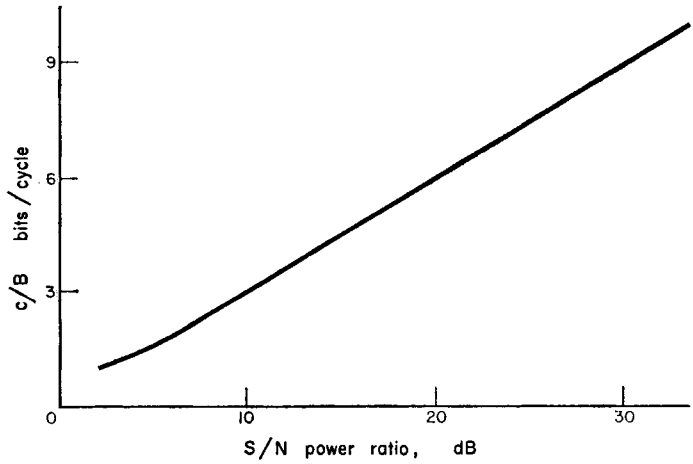


FIG. III.4. Information capacity per unit bandwidth as a function of the signal-to-noise ratio.

The most important problem in the study of transmission is the matching of the characteristics of the message to fit those of the communication channel. The message is characterized by the transmission rate, $R_{message}$, and the transmission channel by its capacity.

The fundamental relation which connects these two characteristics is expressed by the Tuller–Shannon theorem, which may be stated as follows:

"A channel of capacity C can in theory transmit any message which has an information velocity less than or equal to C, irrespective of the natures of the message and the channel".

$$R_{\text{message}} \leqslant C_{\text{channel}}.$$

By way of example, take the standard case of an LF signal transmitted by means of an HF carrier signal. The characteristics of the HF channel are deduced from those relating to the LF channel.

From above we obtain the following inequalities:

$$R_{\text{message}} \leqslant (R_{\max})_{\text{LF}} \leqslant C_{\text{HF}},$$

$$R_{\text{message}} \leqslant \mathscr{B}_{\text{LF}} \log_2 \left(\frac{P}{N}\right)_{\text{LF}} \leqslant \mathscr{B}_{\text{HF}} \log_2 \left[1 + \left(\frac{S}{N}\right)_{\text{HF}}\right].$$

For higher values of the signal-to-noise ratio, the optimum condition takes the form

$$\left(\frac{S}{N}\right)_{\text{LF}} \simeq \left[\left(\frac{S}{N}\right)_{\text{HF}}\right]^{\mathscr{B}_{\text{HF}}/\mathscr{B}_{\text{LF}}}.$$

Contrary to what might have been anticipated, it is possible to transmit the LF message which occupies a frequency band $\mathscr{B}_{\text{LF}}$ by means of a HF channel with a bandwidth $\mathscr{B}_{\text{HF}}$ narrower than that of the LF channel, provided a variation in the signal-to-noise can be tolerated. This is not, however to be recommended, since for a reduction of the bandwidth by half, the signal-to-noise ratio expressed in decibels would have to be doubled, and this would be an extremely high price to pay for the reduced bandwidth.

III.4. APPLICATIONS TO THE CASE OF MULTIPLEX LINKS

In a multiplex link, several transmission channels are grouped together, and the total information is transmitted along a signal path, the communication channel. This path may be provided by the RF carrier wave, or by a telephone cable. In the first case, we have an electromagnetic wave at the carrier frequency, while in the second, a multiplex system with a carrier current. In this book, we shall only be concerned with the electromagnetic wave case.

The essential aim of the calculations in dealing with a multiplex link is to balance the different channels in such a way as to obtain a uniform information rate; this is termed matched coding. The basic symbols are then equally probable and the capacity of the HF transmission channel is

$$C = \mathscr{B}_{\text{HF}} \log_2 \left(1 + \frac{S}{N}\right).$$

The efficiency of the multiplex system is defined by the ratio of the information rate which can be used in the p channels to the capacity of the transmission channel C, or

$$\eta = \frac{pR}{C} = \frac{pf_c \log\left(1 + \frac{S}{N}\right)}{\mathscr{B} \log\left[1 + \left(\frac{S}{N}\right)_{\text{HF}}\right]}.$$

We assume that the frequency band necessary for each channel is constant and equal to f_c, and that the ratios which are fixed by the initial conditions in each channel are identical.

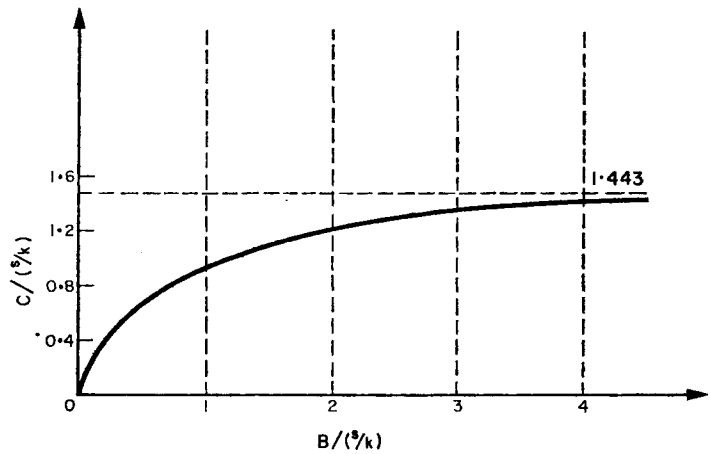


FIG. III.5. Capacity expressed as a function of the bandwidth in normalized units.

The efficiency, so far as information is concerned, is therefore inversely proportional to the capacity of the transmission channel.

$$C = \mathscr{B} \log\left[1 + \left(\frac{S}{N}\right)_{\text{HF}}\right].$$

For a given signal level

$$C = \mathscr{B} \log\left[1 + \frac{S}{k\mathscr{B}}\right]$$

and this is shown plotted in normalized form in Fig. III.5.

When the bandwidth tends to infinity, the capacity reaches a maximum and is equal to

$$C_{\max} = \frac{S}{k} \log_2 e \quad \text{bits/sec.}$$

We introduce the normalized quantities,

$$r = \frac{\mathscr{B}}{pf_c} \quad \text{and} \quad s = \frac{S/N}{(S/N)_{\text{HF}}};$$

r and s represent respectively the spreading of the band and the increase in the signal-to-noise ratio in the multiplex system, compared with a system which has p distinct channels. The expression for the efficiency becomes

$$\eta = \frac{1}{r} \frac{\log S/B}{\log \left[1 + \frac{1}{s}\left(\frac{S}{B}\right)\right]}.$$

The values of r and s are proportional to $\mathscr{B}_{\text{HF}}$. This gives the following useful result: the information efficiency increases when the HF band is narrow for a given low frequency signal-to-noise ratio. The problem, therefore is to get the maximum amount of information into the minimum acceptable bandwidth.

III.5. SOME EXAMPLES OF THE APPLICATION OF THE CONCEPT OF INFORMATION EFFICIENCY

The information efficiency for the various different kinds of ordinary modulation will be compared, assuming that a LF signal can be distinguished when the power ratio, S/N is equal to two, and employing a single channel link ($p = 1$).

Applying the preceding results to the case of amplitude modulation we have

$$r = \frac{\mathscr{B}}{f_c} = 2.$$

The expression $s = \dfrac{(S/N)}{(S/N)_{\text{HF}}}$ is proportional to the ratio of the useful power in the band f_c to the HF power in the sidebands, and inversely proportional to the ratio of the bandwidths. This means that

$$s = \frac{1}{2} \frac{\mathscr{B}_{\text{HF}}}{\mathscr{B}_{\text{LF}}} = 1.$$

For amplitude modulation, there fore, the optimum efficieney is given by

$$(\eta_{\text{AM}})_{opt} = \frac{1}{2} \frac{\log 2}{\log [1 + 2]} = 30 \text{ per cent}.$$

In the case of frequency modulation, the HF bandwidth is

$$\mathscr{B} = 2(m + 1)f_c,$$

and therefore

$$r = 2(m + 1),$$

and the ratio s is equal to $1/3m^2$. The efficiency η_{FM} is therefore of the form

$$(\eta_{FM})_{opt} = \frac{1}{2(m+1)} \frac{\log (S/N)}{\log \left[1 + 3m^2 \frac{S}{N}\right]}.$$

In Fig. III.6 η_{FM} is plotted as a function of m for various values of the ratio S/N (2, 4 and 10).

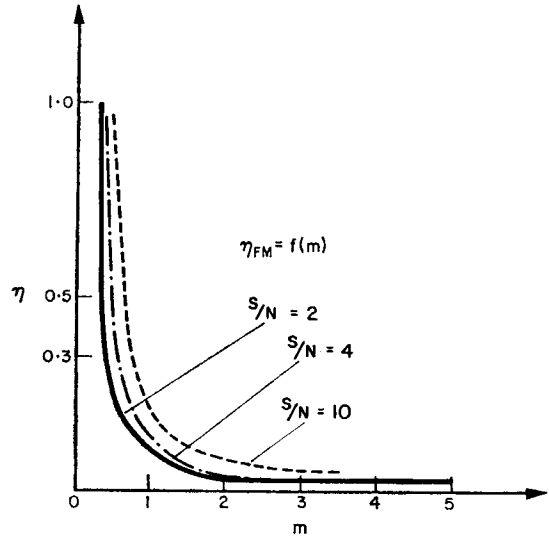


FIG. III.6. Information efficiency in FM as a function of the modulation index.

For a given modulation index m, the improvement obtainable by increasing the ratio S/N is slight, remembering the difficulties which have to be overcome if high values of the ratio S/N are to be achieved.

It is seen that the information ratio is relatively poor when the modulation index is greater than or equal to unity, compared with its value in amplitude modulation.

For very small values of m, we have

$$\mathscr{B}_{HF} = 2f_c$$

or $r = 2$, and the information efficiency becomes equal to that of amplitude modulation.

In the limiting case, $S/N = 2$, the efficiency becomes

$$\eta_{FM} = \frac{\eta_{AM}}{m+1}.$$

As a result, the information efficiency is always less in FM than in AM, whatever may be the value of m. In the case of single sideband modulation we find

$$r = 1 \quad \text{and} \quad s = 1$$

and the efficiency is therefore

$$\eta_{\text{SSB}} = \frac{\log \frac{S}{N}}{\log\left[1 + \frac{S}{N}\right]}.$$

In order to compare this with the other modulation systems, we write $S/N = 2$, and we obtain

$$\eta_{\text{SSB}} = 50 \text{ per cent},$$

which is, by far, the best result from among the three modulation systems considered.

Consider now the information efficiency of a PCM system. The minimum information rate of the message is

$$R = n \log M.$$

If q is the total number of digits, the number of levels, M is equal to 2^q. The rate of transmission n is the minimum frequency of the pulses, $2qF_{\text{max}}$; F_{max} is the maximum modulation frequency.

Thus

$$R_{\text{min}} = 2qF_{\text{max}} \log 2^q$$

$$= 2q^2 F_{\text{max}} \log 2.$$

The minimum bandwidth which is necessary to transmit the coded modulation with q binary digits (bits) is

$$\mathscr{B}_{\text{min}} = qF_{\text{max}},$$

thus the capacity of the channel is

$$C = qF_{\text{max}} \log\left[1 + \frac{S}{N}\right],$$

and the efficiency, therefore, can be written in the form

$$\eta = \frac{R}{C} = \frac{2q \log 2}{\log\left[1 + \frac{S}{N}\right]}.$$

In the Chapter entitled "Radio links for remote-control and telemetry equipment, regarded as multiplex systems" (VI), it is shown that for coded modulation, the signal-to-noise ratio is expressed by

$$S/N = 6p^2$$

in which p is the number of quantization levels:

$$p = 2^{(q-1)}.$$

The efficiency, therefore, may be written as

$$\eta \simeq \frac{2q \log 2}{\log 6 + 2(q-1) \log 2}.$$

For example, taking $q = 7$; η is equal to 0·96. This value is a maximum, as it corresponds to minimum bandwidth. In the majority of cases, the bandwidth accepted is 1·5 $\mathscr{B}_{\min}$, and the efficiency becomes

$$\eta = 0{\cdot}64.$$

From the point of view of information, the PCM system is a most attractive one.

III.6. SENSIBLE USE OF INFORMATION FOR LONG-RANGE LINKS

There are two equally good ways of ensuring the security of a link over a very great distance: either only the minimum amount of information which is strictly necessary is transmitted, or alternatively, a high redundancy is used.

III.6.1. *Limitations on the Information to be Transmitted*

The noise is proportional to the width of the frequency band which is used, and wence the latter is restricted to the absolute minimum; this presumes a knowledge of the signal.

The signal, may, for example, contain within its spectrum components at less than 10 c/s, onto which are superimposed oscillations of about 100 c/s. The main signal at 10 c/s is transmitted with a signal-to-noise ratio of $1/x$, a function of the desired accuracy (x per cent), separately from the superimposed oscillations and along two quite distinct channels. The relative amplitude of the latter oscillations is y per cent, and it is transmitted with a lower signal-to-noise ratio y/x.

The total capacity, therefore, is

$$C = 10 \log \left(1 + \frac{1}{x}\right) + 100 \log \left(1 + \frac{y}{x}\right).$$

For

$$x = 1 \text{ per cent} \quad \text{and} \quad y = 5 \text{ per cent},$$

$$C = 10 \log_2 100 + 100 \log_2 6 = 327 \text{ bits/sec}.$$

Had the signal been transmitted along a single channel, with bandwidth 100 c/s, and a S/N ratio of 100, the necessary capacity would have been

$$C' = 100 \log_2 100 = 670 \text{ bits/sec}.$$

There is thus a saving of information, in the ratio $C'/C = 2$.

III.6.2. *Instruments with a High Redundancy*

We can also conceive of the converse possibility. Consider the case of PCM transmission, in which a word is defined to be the combination of three digits. Allotting two extra digits "1" to each word, a word will only be taken into consideration if there are two digits "1" among the "5" suitably placed. We thus eliminate random fluctuations at the receiver, and at the same time increase the signal-to-noise ratio.

These highly redundant arrangements are extremely interesting, above all where a jammer is liable to interfere with the useful signal. In general terms, the principle behind the arrangement consists of introducing auto-correlation between the symbols which define the useful signal.

These high-redundancy processes will prove to be more particularly applicable to remote-control systems.

III.6.3. *An Example of Detection with Correlation*

The problem is to detect a signal in the presence of noise. We first estimate the useful signal and compare it with the real signal.

Let Γ be the assumed signal, and S and N the real signal and the noise at the input of the detector respectively. After forming the product $\Gamma(S + N) = \Gamma S + \Gamma N$, the signal is passed through a low-pass filter which smooths the information to $\overline{\Gamma S}$ and $\overline{\Gamma N}$, and we finally obtain the signal $\overline{\Gamma S} + \overline{\Gamma N}$. After filtering, the noise N corresponds to that in the low-pass filter band, and the signal-to-noise ratio at the output of the apparatus is

$$\frac{\overline{\Gamma S}}{\overline{\Gamma N}} = \frac{\bar{S}}{\bar{N}}.$$

But, the noise N, which is proportional to the low-pass filter bandwidth, is far weaker and the useful signal S can be extracted.

Figure III.7 is, a block diagram which illustrates this principle. Consider a standard FM receiver to which is added a feed back loop (see Fig. III.8). The frequencies of the supply and of the output of the controlled oscillator are denoted by f_1 and f_2 respectively:

$$f_1 = f_{01} + \Delta f_1 \sin \Omega t,$$

$$f_2 = f_{02} + \Delta f_2 \sin \Omega t.$$

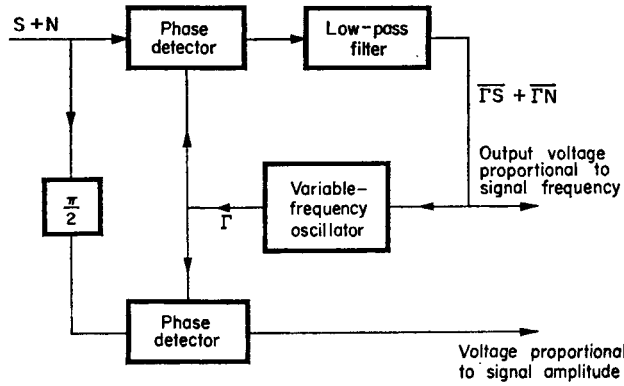


FIG. III.7. Correlation detection.

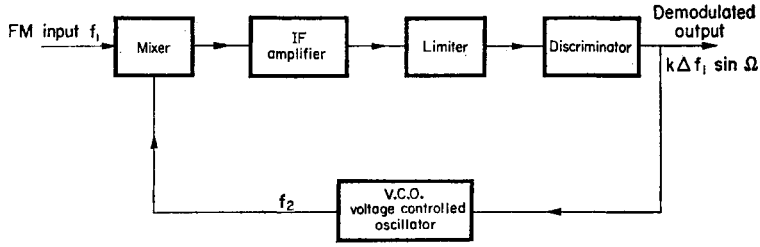


FIG. III.8. FM detection with feedback.

At the output of the mixer stage, the IF frequency is

$$f_{IF} = (f_1 - f_2) = (f_{01} - f_{02}) + (\Delta f_1 - \Delta f_2) \sin \Omega t$$

$$= (f_{IF})_0 + (\Delta f_1 - \Delta f_2) \sin \Omega t.$$

In the case of a standard receiver, the bandwidth necessary at the IF to transmit the modulation is

$$\mathscr{B} = 2(\Delta f_1 + F_{max}).$$

With the feedback system, the bandwidth becomes

$$\mathscr{B}' = 2(\Delta f_1 - \Delta f_2 + F_{\max})$$

and the reduction, therefore, is expressed by the ratio

$$\mu = \frac{\Delta f_1 - \Delta f_2 + F_{\max}}{\Delta f_1 + F_{\max}}$$

or

$$\mu = 1 - \frac{\Delta f_2}{\Delta f_1 + F_{\max}} \simeq 1 - \frac{\Delta f_2}{\Delta f_1}.$$

The influence of the noise, therefore, is reduced in this same ratio.

III.7. CODING

We have seen that a message is defined by the symbols emitted by a source. To code a source of symbols S_i is to substitute combinations of other symbols S_i for them. The conversion from the first to the second constitutes the mechanism of coding. It is obvious that the coding must be reversible, as it must be possible to reconstruct the original message, defined by the symbols S_i, from the symbols S_i' which are received. The operation which is the opposite of coding, at the receiver, constitutes the mechanism of decoding. Four different classes of coding can be distinguished: physical, digital, statistical and secret.

III.7.1. *Physical Coding*

This consists of converting one primary physical quantity into another (the secondary quantity). This operation is performed by a converter. The primary quantity might for example be a pressure, a displacement, or a sound. The secondary quantity is usually of an electrical nature (voltage current, frequency or phase).

The conversion effected by the converter must be linear as far as is possible so that the secondary quantities vary in direct proportion to those of the primary physical phenomenn. This is the basic problem in telemetry. Modulation is a relatively general case of this transformation. This type of coding takes no account of the relations which may exist between the symbols of the source. It is a practical means of transmitting a message which is better adapted than the original to the radio link which is to effect the transmission. The conversion of one physical quantity into another is of an analogue nature; it is a icontinuous transformation. This continuity, which seems *a priori* advantageous, in fact represents the main handicap of the method. The variations in the

signal can in fact be caused not only by those of the primary quantity, as required, but also by subsidiary and random factors such as temperature, voltages, and ageing of the componends; in short, by all the causes of error classified together in Chapter I under the heading of systematic errors. This is the reason why digital coding is often used in instruments for remote-control, telemetry, and distant-signalling; the mode of operation and the advantages of this type of coding are now considered.

III.7.2. *Digital Coding*

The continuous message is quantized into M distinct levels, as we have already seen.

The maximum number of coded combinations which it is possible to define with a number-system of base b is b^n, where n is the number of binary digits (bits). As a result, if one of the $M = b^n$ possible combinations is allotted to each level, a number $1 < N < M$ will be expressed in this system in the form

$$N = a_n b^{n-1} + a_{n-1}^{b-2} + \cdots + a_2 b^1 + a_1 b^0.$$

With each digit a_k is associated a weighting factor $b^{(k-1)}$. The number N is the weighted sum of the digits. As is seen in the appendix to this chapter, the most suitable system is the binary system ($b = 2$).

The quantization error is equal to

$$\varepsilon = 1/M = b^{-n}$$

or in the binary system

$$\varepsilon = 2^{-n}.$$

It is the number of binary digits in the code, n, which determines its accuracy. In the natural binary system, the number 23 is written

$$16 + 4 + 2 + 1$$

or

$$1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$$

or

$$1\ 0\ 1\ 1\ 1.$$

In a weighed code like the natural binary system, the digits 0 and 1 are weighted according to their positions. The weights are set in decreasing order from left to right. There are, however, unweighted codes in which only the grouping of the digits is of importance, and this is the case of the reflected binary code. Two consecutive levels then differ only by a single binary digit.

In a reflected binary code with n digits, symmetry properties appear with respect to the axes which lie between the levels:

$$(2^{(n-1)} - 1) \text{ and } 2^{n-1}$$
$$(2^{(n-2)} - 1) \text{ and } 2^{n-2}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(2^2 - 1) \text{ and } 2^2$$
$$(2^1 - 1) \text{ and } 2^1$$

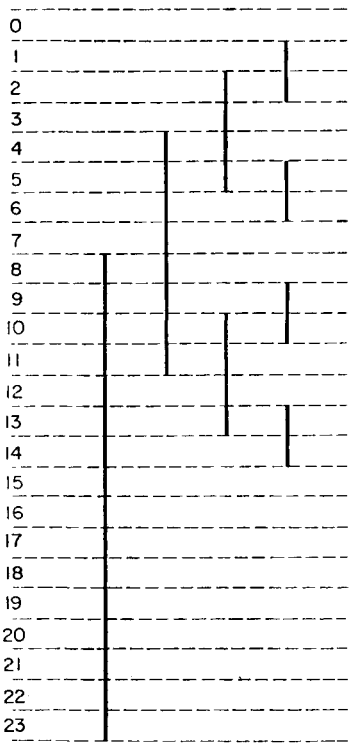


FIG. III.9. Construction of the reflected binary code. Example: 9 = 1101.

This code can easily be constructed by means of Fig. III.9. The fact that only one digit varies at a time on changing from one level to the next restricts the possibilities of error in coding. Let us take a concrete example: assume that the level in question lies between 7 and 8. In the natural binary code, the level 7 is represented by 0 1 1 1 and the level 8 by 1 0 0 0. For the level 7·5 therefore, the coder would oscillate between the two alternatives which represent two complementary states (digit 0 replaced by digit 1 and vice-versa).

In the reflected binary code, on the other hand, the level 7 is written 0 1 0 0 and the level 8, 1 1 0 0. For 7·5, the coder would give 0 1 0 0 or 1 1 0 0; the possible error is therefore of a single digit only, which corresponds to one quantization interval. It is only necessary to choose a number of levels n high enough to minimise the error to within the desired value.

The fact that the code is unweighted does, however, result in a more complicated decoding system; it is, in fact, necessary to convert the reflected binary into natural binary code at the receiver, before finally decoding into the original quantity.

III.7.3. *Statistical Coding*

In this coding category the probabilities of the various symbols appertaining to the message are taken into consideration. The message is then divided into groups of symbols with different probabilities. The shortest code groups will be allotted to the highest probability. In these procedures, it is often useful to store the data in a memory. As the transmission time of messages of the same length may be spread out to a greater or lesser extent, delays have to be introduced into the transmission of certain sequences in order to obtain a uniform velocity. This is the case which is most favourable to the matching of the message source to the channel capacity.

We know that the capacity of a channel can only be approached ($C = \mathscr{B} \log_2(1 + S/N)$); it can only be reached if the signal possesses the characteristics of white noise. The problem therefore, is to ensure an efficient transmission with a reduced capacity. It has been seen that the information rate R is expressed in the form:

$$R = n \log_2 M = nH \quad \text{bits/sec}$$

in which M is the total number of distinct levels (or symbols).

If the symbols S_i appear at a rate of n per second, and are coded into a succession of symbols S_i' at a rate of N' per second, we have:

$$R = R'$$

or

$$nH = n'H'.$$

H and H' respectively are the entropies which correspond to the distinct levels. M and M' But in most cases there is a correlation between the different symbols, and this reduces the number of similar sequences of a given length; thus

$$R' > R.$$

It is therefore possible to reduce the capacity of the transmission channel in the ratio R/R' (the coding is impossible if $R' < r$). For example, the orders

"dive" and "climb", encounterod in remote control, are mutually exclusive, and the correlation which exists between them is exprossed by their simultaneous incompatibility.

It has also been seen that if the symbols are independent, but have different probabilities, the value of the entropy is:

$$H = -\sum_i P(i) \log P(i).$$

When, on the other hand the symbols are dependent upon one another, when that is to say the probability of the presence of a state "i" is conditional upon the presence of a state "j", we have:

$$H = -\sum_i p(i) \sum_j P_i(j) \log P_i(j)$$

$$= -\sum_j p(j) \sum_i P_j(i) \log P_j(i).$$

Shannon has shown how a code can be constructed, using only the digits 1 and 0, which takes into account the following two fundamental principles: the shortest code group should be allotted to the most frequent sequence, and these groups should be distinguished by some other characteristic than separation, or some provision should be made to prevent any code group from being duplicated.

Symbol denoted by	Probability	Code
No. 1	1/2	0
2	1/4	1 0
3	1/8	1 1 0
4	1/16	1 1 1 0
5	1/32	1 1 1 1 0
6	1/64	1 1 1 1 1 0

FIG. III.10. An example of a Shannon–Fano code.

By way of example, consider a possible code distribution shown in Fig. III.10 (the Shannon–Fano code) in which the successive symbols are placed in order of decreasing probability. Each group of the code consists of a series of digits 1 followed by the digit 0; the presence of the latter is to be regarded as symptomatic of the end of a code group.

In the case of Fig. III.10 the value of the entropy before coding is found to be

$$H = 1{\cdot}7 \text{ bits}.$$

After coding, provided the number of symbols is great enough, the mathematical expectations of "1" and "0" tend to become equal to 1, so that the probabilities of "0" and "1" are equal to $\frac{1}{2}$, and the entropy of the system becomes

$$H' = \tfrac{1}{2}\log\tfrac{1}{2} + \tfrac{1}{2}\log\tfrac{1}{2} = 1.$$

The improvement factor achieved with this type of coding is 1·7; while preserving the same transmission rate, the capacity of the transmission channel can be reduced in the ratio of 1·7 : 1.

Symbol denoted by	Probability	Code
No. 1	1/4	0 0
2	1/4	0 1
3	1/8	1 0 0
4	1/8	1 0 1
5	1/16	1 1 0 0
6	1/16	1 1 0 1
7	1/32	1 1 1 0 0

FIG. III.11. Another example of a Shannon–Fano code.

As a second example, consider another coding of the Shannon–Fano type shown in Fig. III.11.

The number of digits which correspond to any one symbol varies as a function of the probability allotted to it, and the operation of differentiating between symbols is effected by aknowledge of the digit on the right.

III.7.4. *Secret Coding*

This is the problem which appears mainly when a tactical missile is to be remote-controlled, and must not be interfered with by enemy transmissions. A series of "keys" are prepared which will unlock the receiver. Only the "friendly" signallers know the secret of the combinations, and can arrive at the succession of steps which alone lead to eventual detection.

Coding here consists in having at one's disposal a whole range of combinations in cascade, which can be rapidly changed, so that the enemy will be ignorant of which one is in use at any given time.

This is a problem of a different nature from the one of methodical coding studied earlier within the framework of information theory. These systems are necessarily so highly redundant that a perturbation of an element does not result in disturbance of the whole.

The receiver of the remote control system lies within the missile itself, so that the space at one's disposal is limited; a choice of coding method which avoids an unneccessary multiplicity of circuits is therefore necessary. Sometimes a more complicated coding method is used in order to allow the decoding to be performed with the minimum of circuitry.

One efficient means of secret coding consists in transforming the original signals with the aid of supplementary signals. To modify the message and to make it intelligible, the entropy of the supplementary signal must be greater than the quantity of information contained in the original message. It should be impossible to extract the useful information contained in the message without a knowledge of the auxiliary which has been superimposed on the original. As a result there will be correct reception only if the jammed signal and the supplementary signals are handled simultaneously at the transmitter and the receiver. This synchronism is possible by establishing beforehand a programme of transformations for the signals.

III.8. METHODS OF CODING AND DECODING

We shall now consider briefly a few practical ways of solving the problems set by coding and decoding, which are used in industry.

III.8.1. *Digital Coding*

This transformation can be carried out in several different ways, by employing either photo-sensitive elements, capacitative effects, the properties of the cathode ray tube, or purely electronic circuits.

Consider a disc which is free to rotate about its axis and which is provided with slits through which the light from a lamp can pass. A diaphragm is placed facing the disc to prevent the photocells from being affected by adjacent light beams (see Fig. III.12).

For a binary code with n binary digits, for example, we use n photo-sensitive elements. The positioning of the slits on the disc corresponds to the coding which has been selected (natural or reflected binary). The presence or absence of a slit corresponds to the digit 1 or 0. This arrangement is ideal for the transmission of an angle. To transmit a voltage, the coding disc is

attached to the needle of a voltmeter. In general the electrical or mechanical quantity which is to be coded determines the angular position of this disc.

The photo-sensitive elements are often photo-diodes or photo-transistors, which provide relatively large currents, which can be used directly. Thus all the digits of the code are given simultaneously set out in a "parallel" form.

Another type of parallel coder has been studied by the CNET*—known as a multiple-frequency coder. It is arranged that the information is applied to the coder in the form of a series of pulses of variable duration. When a pulse appears, n oscillators (in the case of a code with binary digits) are

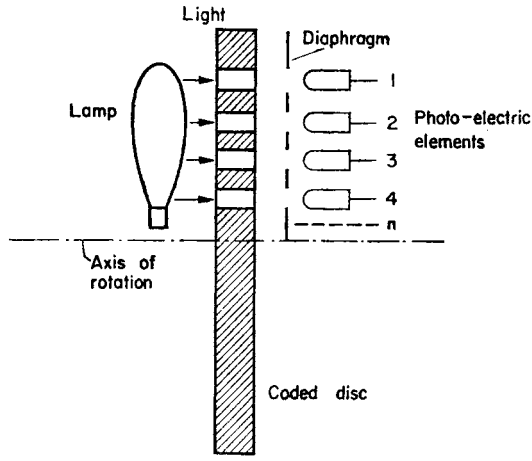


FIG. III.12. Coding disc.

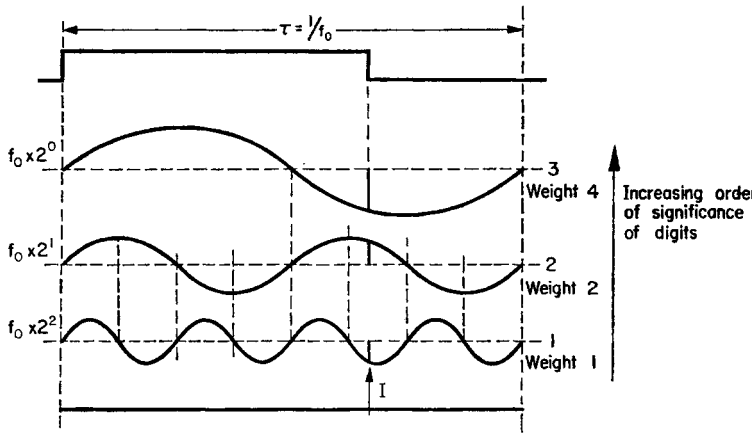


FIG. III.13. Operation of a multiple frequency coder.

* Centre National d'Etude des Télécommunications (France).

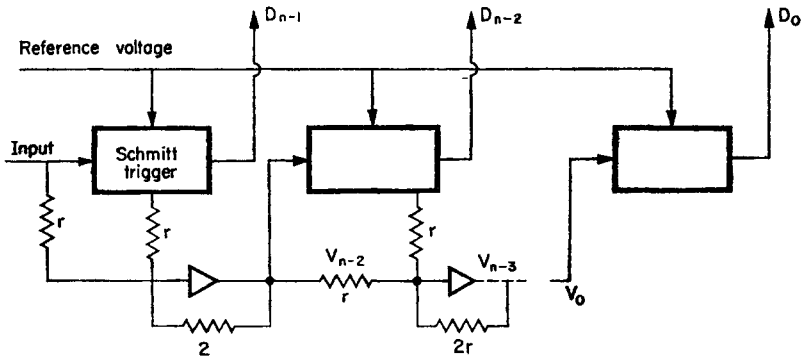


FIG. III.14. Coding by means of stages in cascade (Smith).

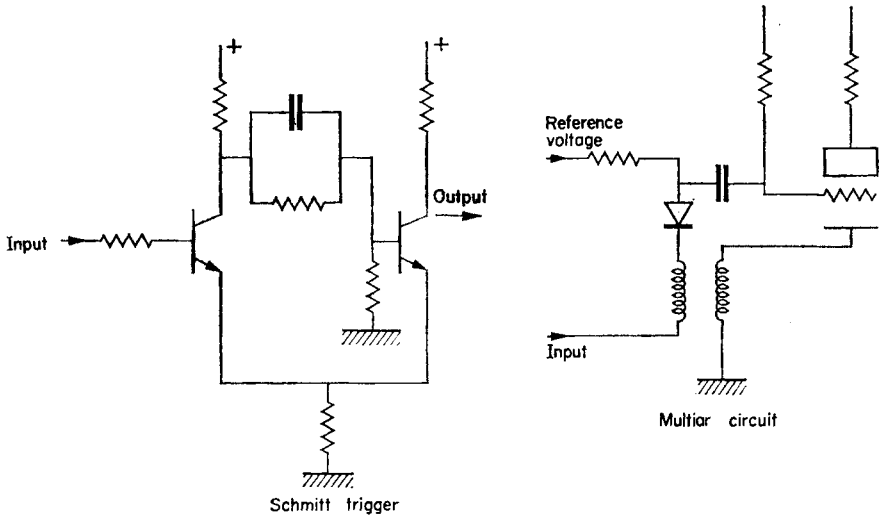


FIG. III.15. Comparison circuits.

triggered off. The frequencies of these n oscillators are equal to $2^0 f_0$, $2^1 f_0$, $2^2 f_0$, ..., $2^{n-1} f_0$ respectively, with $f_0 = 1/\tau$ where τ is the maximum pulse-length. A pulse is produced by differentiation of the falling edge of the input pulse and is compared for coincidence with the n oscillator states. In the case shown in Fig. III.13, when the pulse 1 to be coded is compared with the sinusoidal signals (lines 1, 2 and 3) then we obtain the code 1 0 1. We notice that the greatest weight is associated with the lowest frequency.

When remote-control instructions are to be transmitted as pulses, the simplest means of coding is to open or close a relay. Suppose there are four remote-control commands to transmit; a number is allotted to each and it is

only necessary to translate the presence or absence of this order into the binary system to achieve a means of coding the order.

A very interesting coding system is that due to B.D.Smith, employing stages in cascade (see Fig. III.14). The input voltage is compared with a reference voltage, judiciously chosen for each stage. If the comparison reveals that the input signal is higher than or equal to the reference potential, the comparison circuit indicates "1", and a subtractor deducts the reference voltage from the input voltage; otherwise, it registers "0", and the input signal passes on to the next stage unaltered.

By using separate references for each stage, the output of each subtractor is multiplied by two, and each reference voltage is made equal to $2^{n-1}\varepsilon$ volts, where ε is the quantisation level and n is the number of stages used.

The comparison circuit is a Schmitt trigger, or a Multiar circuit (see Fig. III.15) which gives an output signal of 0 volts when the input voltage is less than the reference R (and $-R$ in the opposite sense). The gain in each amplifier is equal to two; if V_{j-1} is the output voltage of the j-th amplifier, we have:

$$V_{j-1} = 2(V_j - RD_j),$$

where

$$D_j = \begin{cases} 0 & \text{if } V_j < R \\ 1 & \text{if } V_j \geqslant R \end{cases}.$$

Consider the example of a five-stage coder, calculated to work with an input which may vary from zero to 31 volts.

$$M = 2^5 = 32.$$

The quantization error, ε, is therefore 1/volt and the reference potential is $2^{(5-1)} = 16$ volts.

For an input potential of 23 volts,

$$\begin{aligned} V_4 &= 23 && \text{implies} && D_4 = 1 \\ V_3 &= 2(23 - 16) = 14 && \text{implies} && D_3 = 0 \\ V_2 &= 2 \times 14 = 28 && \text{implies} && D_2 = 1 \\ V_1 &= 2(28 - 16) = 24 && \text{implies} && D_1 = 1 \\ V_0 &= 2(24 - 16) = 16 && \text{implies} && D_0 = 1. \end{aligned}$$

III.8.2. *The Parallel–Series Code Converter*

Very often, the information is first coded in parallel form, and then transmitted in series form. To perform this transformation, code transformers are required. By distributing the information cyclically, it is possible to transmit several sets of information along the same channel.

There are two kinds of code converter or adaptor: "delay-line" systems and bistable systems. In the first (see Fig. III.16), clock the signal is applied to the input of the delay-line, and simultaneously, the parallel code is fed into coincidence circuits. The signal appears in series form at the output of the coincidence circuits.

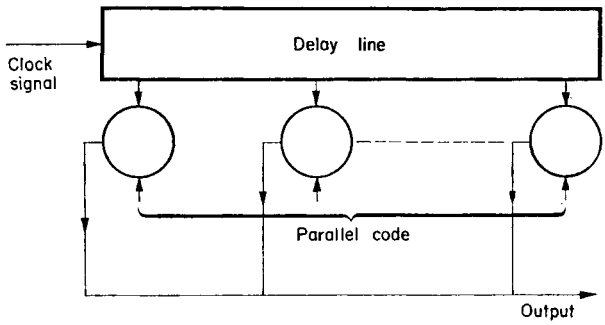


FIG. III.16. Parallel-series converter using a delay line.

We must ensure that the frequency of the clock signal and the total delay introduced by the line are compatible with the duration of the interval over which the pulses of the parallel code are applied.

Another system with which carries out a parallel–series code transformation depends upon the use of a series of bistable elements, and this is the procedure in most common use. Often a small magnetic core with a rectangular hysteresis loop is used as the bistable element (see Fig. III.17).

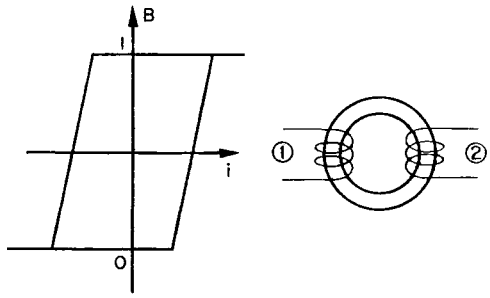


FIG. III.17. Cores with a rectangular hysteresis loop.

The magnetic material has two stable states which correspond to the saturation flux densities $+B_s$ and $-B_s$, which are taken as the states "1" and "0" respectively. A voltage is induced across the winding 2 only when a change of state occurs.

Suppose that by means of a diode, connected appropriately in the second winding, we arrange that a usable voltage is induced in winding 2 only

when the change of state is from state 1 to state 0; thus, the change from 0 to 0, from 1 to 1, or from 0 to 1 yields the digit "0", whereas changes from 1 to 0 yields "1".

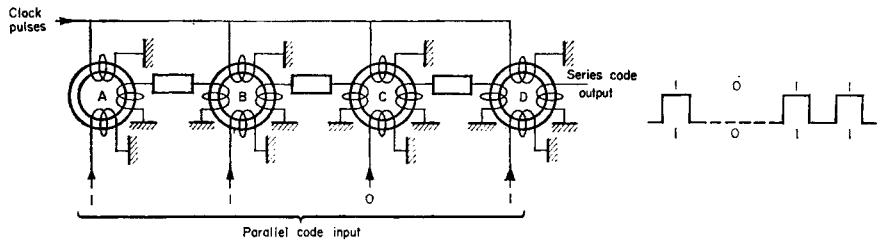


FIG. III.18. Parallel-series converter with bistable elements.

Consider an actual example: a message coded in the binary system in parallel form, the number 1101 say, is required in series form. The number is applied to four bistable elements as shown in Fig. III.18.

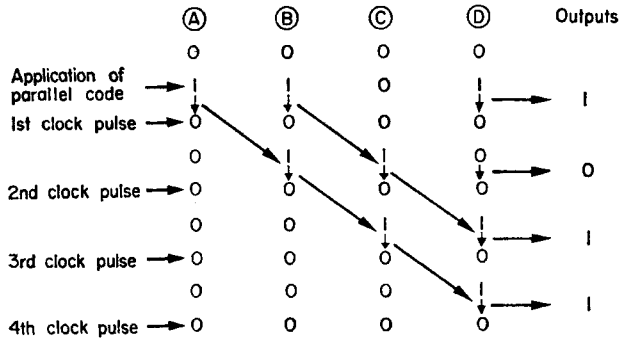


FIG. III.19. Sequence of operations in a parallel-series converter using bistable elements.

The circuits *A*, *B*, *D* change from state 0 to state 1, while circuit *C* remains in state 0. The states of the circuits are 1, 1, 0 and 1 respectively. A clock pulse resets the circuits to zero, and those which pass from state 1 to state 0 induce a digit in the subsequent circuit. The procedure can be followed easily with the aid of Fig. III.19.

At each clock pulse, the system clears itself step-by-step of its digits. Finally, all the elements are reset to their initial state 0, and the whole process can begin afresh.

It should be noticed that the clock signal frequency must be n times greater than the repetition frequency of the code groups if n is the number of binary digits.

The digits, when they are given in parallel form, appear backwards (the reverse of 1101 is 1011). It is necessary therefore to connect the entries in the reverse order to obtain the code group in its direct form.

III.8.3. *Series–Parallel Code Converter*

The transformation of a series code into a parallel code is a very simple operation; it is only necessary to apply the synchronization pulse to a delay line, and to form coincidences between the signal in its series form and the delayed pulses obtained from the line, as shown in Fig. III.20.

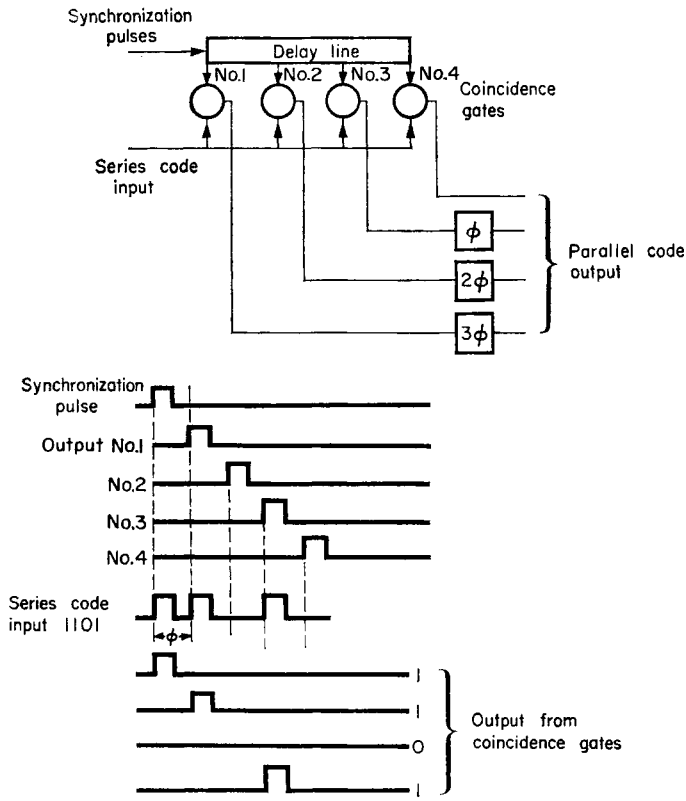


FIG. III.20. Series-parallel code converter using a delay line.

At the output from the coincidence circuits, we obtain pulses whose phases are varied to a greater or lesser extent according to the weights associated with them. Thus the outputs 3, 2 and 1 correspond to phase

changes of φ, 2φ and 3φ respectively. For a code with n binary digits, $(n-1)$ phase shifts are necessary.

When biased transistors or valves are used as bistable devices, the reasoning is identical.

III.8.4. *Digital-analogue Decoding*

This process is the inverse of coding. Starting with a number expressed in the binary system, an analogue quantity is produced (current, voltage, or displacement for example), proportional to the quantised value of the code, corresponding to this number, or more exactly to this coded combination. The device which performs this transformation is styled a "decoder". Decoder inputs are usually in parallel code, and the analogue quantity is usually a current.

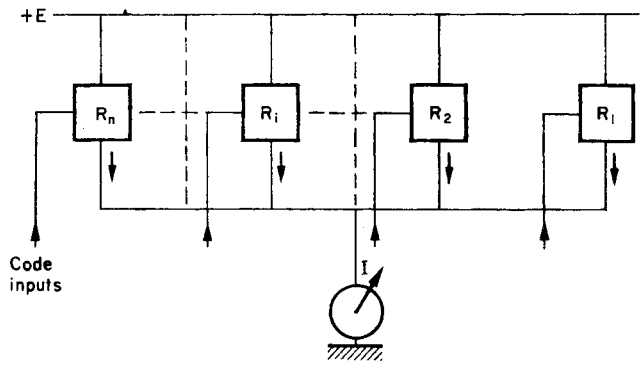


FIG. III.21. A general arrangement for decoding.

A source of direct current, at a constant voltage E, is connected to n variable resistance devices $R_1, R_2, \ldots, R_n$, which correspond to the n binary digits of the code. An element R_i has a finite resistance in the presence of a digit "1", and is effectively infinite in the presence of "0". The values of the resistance $R_1, R_2, \ldots, R_n$ are proportional to the respective weights $2^0, 2^1, 2^2, \ldots, 2^n$ associated with each moment. The general arrangement is shown in Fig. III.21.

The elements R_i may consist of circuits which display an infinite resistance in their unenergized positions. When a digit "1" is applied to the control winding, the contacts close and the voltage E is connected across the ends of the resistance $R_i = 2^i R$.

Consider a code group with four binary-digits 1101, which is to be translated into analogue form. The states of the elements R_1, R_2, R_3 and R_4 are, respectively:

$$R_1 \begin{matrix}\nearrow \infty \\ \searrow R\end{matrix} \quad R_2 \begin{matrix}\nearrow \infty \\ \searrow 2R\end{matrix} \quad R_3 \begin{matrix}\nearrow \infty \\ \searrow 4R\end{matrix} \quad R_4 \begin{matrix}\nearrow \infty \\ \searrow 8R\end{matrix}.$$

The current I which passes through the milliammeter is $I = E/\varrho$, in which ϱ is the equivalent resistance of the bank of resistors in parallel, namely

$$\frac{1}{\varrho} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \frac{1}{R_4}.$$

The maximum current, I_{max}, is given by

$$I_{max} = \frac{E}{\varrho_{min}} = \frac{E}{R}\left(\frac{1}{8} + \frac{1}{4} + \frac{1}{2} + 1\right)$$

or

$$I_{max} = \frac{15E}{8R}.$$

The minimum current is zero.

For the combination 1101,

$$\frac{1}{\varrho} = \left(\frac{1}{8} + \frac{1}{4} + 1\right)\frac{1}{R} = \frac{11}{8R}.$$

The current I therefore has the value $I = 11/15\, I_{max}$.

Diodes can equally well be used for the switching elements R_i, and would present finite or infinite impedances according to their biasing. Again transistorized elements can be used; these would normally be cut off by means of Zener diodes, and would conduct only if a pulse in the right sense was applied to their bases. In both cases, the milliammeter measures the sum of the currents which flow through the switching circuits.

There are other decoding devices, such as the coincidence decoder, in which the analogue converter again works by forming the weighted sum of the code pulses it receives.

The RC circuit decoder is based on the discharge characteristic of a capacitor, and on the fact that between two code pulses the charge in the capacitor reduces by half.

APPENDIX — REASONS FOR CHOOSING THE BINARY SYSTEM

Every measurement of a physical quantity can be reduced to the determination of an amplitude which is expressible in the form of a number (1 volt, 12 volts ...); the choice of the base b of the notation is thus of prime importance.

Any number A can be written in the form $A = a_1 + a_2 b + a_3 b^2 + \cdots + a_n b^{(n-1)}$. The power n attached to the base represents the relative weight of the corresponding coefficient.

The number of combinations which can be defined is

$$M = b^n.$$

and the quantity of information, therefore, is

$$Q = n \log b = \log M.$$

The number of circuits will be a minimum if the product nb is a minimum, the quantity of information remaining constant. We write

$$y = nb = \frac{Qb}{\log b},$$

and differentiate:

$$dy/db = Q\left[\frac{1}{\log b} - \frac{b}{(\log b)^2} \times \frac{1}{b}\right] = \frac{Q}{\log b}\left[1 - \frac{1}{\log b}\right].$$

dy/db is zero when $\log b = 1$, or when $b = e = 2{\cdot}7$, and the closest whole number is therefore 3. The function y has a very flat minimum, and for $b = 2$, we find $b/\log b = 2{\cdot}8$, whereas for $b = 3$, we have $b/\log b = 2{\cdot}7$. The binary system is more practical than the trinary, and for this reason the binary system of counting is usually employed whenever the system permits it.

CHAPTER IV

LIMITATIONS ON INFORMATION

WE HAVE seen that the maximum rate at which information can be transmitted along a communication channel is limited by the disturbances which naturally exist in all transmission. It is these disturbances which form the subject of this chapter which is divided into four parts—the study of the phenomenon of noise, jamming, systematic errors, and finally, distortions.

It is difficult in a single chapter to cover every aspect of the problem of perturbations since they are in a sense infinite in number and the origins of some of them are still imperfectly understood.

When we replace these complex phenomena by parameters such as noise factor and measurement threshold we are expressing in concrete form the limits to which a calculation of the maximum transmission capability is subject.

IV.1. THE PHENOMENON OF NOISE

Noise is to be regarded as a parasitic signal, independent of the useful signal onto which it is superimposed. In this section, the different kinds of noise inherent in the components employed and internal and external kinds of noise are distinguished. Finally the signal-to-noise ratios in the various types of modulation are evaluated to allow a comparison.

IV.1.1. *Noise Inherent in the Components Employed*

IV.1.1.1. *The Johnson Effect*

The thermal motion of the electrons in a metal obeys the laws of thermodynamics. The free electrons in the conductor constitute what is known as an "electron cloud"; they are in continual agitation, and their movements become more unruly as the ambient temperature is increased. The presence of this agitation in a resistance R results in a fluctuating e.m.f., $e(t)$.

If $\overline{e_f^2}$ is the spectral density of $e(t)$,

$$\overline{e_f^2} = 4KRT$$

in which K is Boltzmann's constant
R is the resistance in question
T is the absolute temperature.

When the impedance is not simply resistive but is complex, $Z = X + jY$, the resistance is replaced by the real part of the impedance, and the spectral density of the noise is then

$$\overline{e_f'^2} = 4KXT. \tag{IV.1}$$

In practice, the e.m.f. due to thermal fluctuations in a frequency interval $(f_1 - f_2)$ is given by

$$E = \sqrt{4KRT(f_1 - f_2)} \tag{IV.2}$$

$$K = 1{\cdot}381 \times 10^{-23} \text{ joules/°K}$$

$$T = 273 + t\,°\text{C}.$$

The thermal e.m.f. developed in the grid circuit of the first stage of a receiver is of the order of one microvolt.

The resistance R is usually in fact the input impedance of a receiver, that is, of a valve or a transistor. The input impedance of a transistor is lower than that of a valve, and it would seem, therefore, that a transistorized amplifier should be better as far as the Johnson effect is concerned. However, there are other inconvenient features which detract from this advantage.

IV.1.1.2. *The Shot Effect*

This is an effect which is present both in valve amplifiers and in transistor amplifiers.

When a direct current is applied to the input of a linear amplifier, the voltage at the output is found to fluctuate on a microscopic scale; this is the shot effect. It should be noticed that the magnitude of this effect increases with increase of the heater voltage of the valves. The noise e.m.f. consists of a series of extremely short pulses, caused by the impacts of the electrons emitted by the cathode on the anode. In the case of transistors, this noise is the result of the current fluctuations produced by the diffusion and recombination of electrons and holes in the semiconductor. These impacts and recombinations follow a probability law, as in all randomly varying phenomena, that is described by a Poisson distribution. We reach the following result: if i is the steady value of the current applied to the input of the amplifier, e the charge of the electron, and $\mathscr{B}$ the width of the frequency band, the mean square value of the fluctuating noise current, $I(t)$, is

$$\overline{I^2} = 2ei\mathscr{B}. \tag{IV.3}$$

A practical way of thinking of the effect of noise is to regard it as a current source in the case of transistors, and a voltage source in a valve amplifier.

The noise e.m.f. produced in the input circuit of the first stage of the amplifier can be determined from the relation:

$$E = \sqrt{(3{\cdot}18 \times 10^{-19} IZ(f_1 - f_2)}, \tag{IV.4}$$

in which I is the electron current in the valve
Z is the impedance of the anode circuit at resonance, which we assume to be constant over the band $(f_1 - f_2)$.

In the preceding discussion, the effect of space charge has not been taken into account. In practice, the e.m.f. produced by the shot effect can be taken as the order of a microvolt, referred to the grid of the first valve.

IV.1.1.3. *Noise in Semiconductors*[26]

Apart from the noise due to the ordinary fluctuations which have just been described, and which are also encountered in semiconductors, there is a special kind of noise peculiar to the latter. It is also found in some types of resistor, carbon microphones, and in semiconductor elements in general. This noise is caused by:

(1) The effect of the thermal fluctuations on the recombination velocities; this type of noise can be produced synthetically by placing two generators between the base-emitter and the base-collector junctions.

(2) The fact that the leakage current I_f at the base-collector junction, biased in the opposite sense, is the difference between the quiescent current I_{CB_0} and the saturation current, I_s.

$$I_f = I_{CB_0} - I_s.$$

This leakage current increases with the potential, V_{CB}, applied between the collector and the base, whereas I_s remains to all intents and purposes constant. It should be noticed that the noise increases with the value of the leakage current.

The preceding results can be summarized quantitatively by expressing the mean square noise e.m.f. in the form

$$\overline{e_s^2} = KVRf^{-1}\mathscr{B}, \tag{IV.5}$$

in which K is a coefficient, which depends upon the semi-conductor employed
V is the steady potential across the terminals of the semiconductor.
R is the resistance as measured with an ohmmeter
$\mathscr{B}$ is the frequency bandwidth
f is the working frequency.

The noise potential will therefore be reduced by using small biases and low resistances R.

IV.1.1.4. *The Noise Factor*

The best way of indicating how much noise an amplifier produces (regardless of type) is to define a "noise factor" for it. This factor is the ratio of the total noise power in the load to the power of the noise which has its origin solely in the thermal agitation in the source resistance. The ideal value of the noise factor is unity, and the closer this factor is to unity, the more closely does the amplifier approach the ideal.

Consider a succession of quadripoles (or four-terminal networks) connected in a chain, with gains $G_1, G_2, \ldots, G_n$ and noise factors $F_1, F_2, \ldots, F$, respectively (see Fig. IV.1).

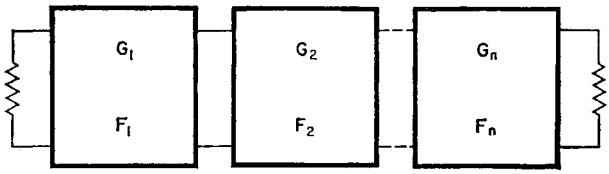


FIG. IV.1. A series of n quadripoles, gain G_n and noise factor F_n.

The total gain, G, of the whole combination is $G = G_1 \cdot G_2 \cdot G_3 \cdot \ldots \cdot G_n$. The overall noise factor, F, of the whole chain has the value

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots + \frac{F_n - 1}{G_1 G_2 \ldots G_{n-1}}. \qquad \text{(IV.6)}$$

When each quadripole corresponds to one stage of the amplifier, the term in $(F_2 - 1)/G_1$ is seen to be negligible, as are thus all the succeeding terms; and the overall noise factor is equal to that of the first stage. In other words, the noise is, so to speak, localized at the input stage.

If, on the other hand, two families of quadripoles are considered in which the first represents the mixer and the second the I.F. stage, we have

$$F = F_1 + \frac{F_2 - 1}{G_1}$$

in which G_1 is now the gain of the mixer. G_1, however, is too small for $(F_2 - 1)/G_1$ to be regarded as negligible.

The noise factor is a criterion of quality which it is often convenient to employ when comparing two similar elements. Thus, for example, for input resistances smaller then 1000 Ω, the noise factor of transistors is lower than that of radio valves. For an input resistance of 1000 Ω, the noise factors of a vacuum tube and a transistor are of the orders of 8 db and 4 db respectively.

The noise factor of a transistor is a function of the frequency employed, as we show in Figs. IV.2, IV.3 and IV.4. From Fig. IV.2 it can be seen that the noise in the semiconductor itself varies predominantly as $1/f$ in part (A) of the curve, whereas in part (B), the noise is independent of the frequency. When the frequency increases, the gain of the transistor decreases and the noise increases. As a general rule, the noise begins to increase for $f \geqslant f_a/10$ when f_a is the cut-off frequency.

IV.1.2. *Atmospheric Noise*

Under this heading, we shall collect together a variety of phenomena, the causes of which are often known, but which can only with difficulty be expressed in a quantitative manner. It is particularly important to take these into account with missiles travelling at very high altitudes, or designed for travel through interplanetary space.

IV.1.2.1. *Environmental Noise*

This noise of is meteorological origin, and affects the propagation of radio waves in the atmosphere. The air is agitated in summer by convection currents as the earth heats up, and this imparts a certain heterogeneity to the ambient medium.

The atmospheric turbulence caused by depressions and winds is also a source of zones which may deflect electromagnetic radiation.

The resulting perturbed medium is relatively homogeneous, but certain atmospheric layers have different water-vapour densities, and the saturated vapour pressure varies with the altitude. A modification of the refractive index of the atmosphere results, and this disturbs wave propagation.

In more general terms, tropospheric disturbances are caused by variations in the refractive index, and are particularly noticeable at relatively high frequencies. An expression which allows the refractive index to be calculated as a function of the temperature T in °K, the pressure P in mb, and the saturated vapour pressure e_s in mb is commonly given in the form

$$(n - 1)\, 10^6 = \frac{79}{T}\left(p - e_s + \frac{4800}{T}\, e_s\right).$$

The refraction error is then

$$\Delta E = \cot E \cdot \frac{1}{H} \int_0^H (n_0 - n)\, \mathrm{d}h$$

in which E is the site angle, n_0 the refractive index at ground level, n the refractive index at an altitude h and H the altitude of the missile.

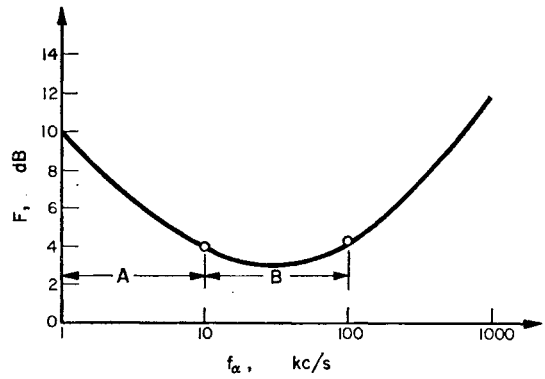


FIG. IV.2. Variation of the noise factor of a *p–n–p* transistor as a function of the cut-off frequency f_α.

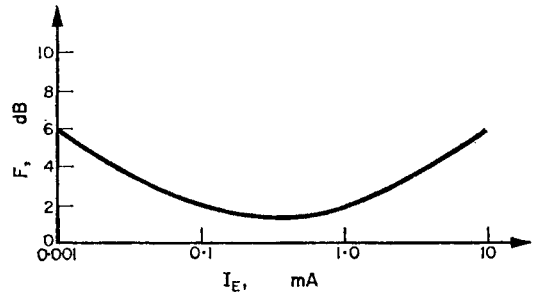


FIG. III.3. Variation of the noise factor of a transistor as a function of the emitter current (common base connection).

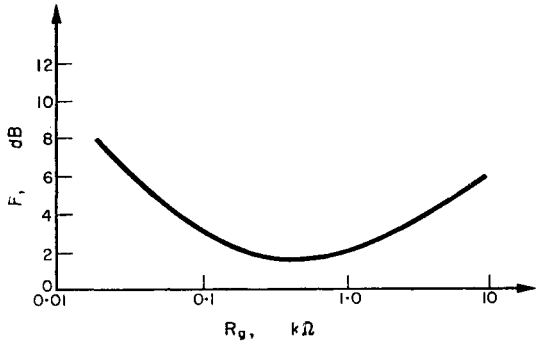


FIG. IV.4. Variation of the noise factor of a transistor as a function of the resistance of the source (common base connection).

For great altitudes,

$$\Delta E \simeq (n_0 - 1) \cot E \quad \mu\text{radians}.$$

Under the same conditions, the distance error is

$$\Delta R = \operatorname{cosec} E \int_0^H (n - 1)\, dh.$$

In the ionosphere, the refractive index becomes

$$n = \sqrt{\left(1 - \frac{80N_e}{f^2}\right)}$$

in which N_e is the electron density per cubic metre, and f is the frequency in question.

Atmospheric fluctuations give rise to noise of a rather special nature which is related to the dynamic properties of the atmosphere, such as changes of pressure, temperature, or humidity. The presence of clouds is the predominant effect in the troposphere, whereas variations in the electron density dominate in the ionosphere.

Atmospheric noise is appreciable in radio transmission at frequencies higher than 500 Mc/s. For frequencies of the order of 1000 Mc/s, the polarization of the wave is affected in the ionosphere, by an unpredictable amount.

IV.1.2.2. *Cosmic Noise*

This kind of noise is the result of ionization in the upper atmosphere; it includes solar noise, which is related to sunspots, and these are more or less active according to the year. This noise may therefore be ten times greater during a period of widespread solar eruptions than during a quieter period.

The noise power density of these sources tends to increase when the frequency increases. For an omnidirectional receiver aerial, localized within the earth's orbit, the cosmic noise at 250 Mc/s would be ten times less than the noise in the ordinary kinds of receiver in present use. In ultra-sensitive equipment such as masers and parametric amplifiers, the receiver noise-level is reduced one hundredfold and as a result, the effect of cosmic noise on the clarity of transmission must be taken into account.

In radio astronomy, it is found that noise is more intense in the directions of the nebulae and of the sun. It is to be remembered that the radiation of radio waves by the sun, as detected by a terrestrial receiver, is on the average 10^{-22} watts/cycle m^2 of the surface illuminated.[44]

It is essential therefore to take these quantities into account when an electromagnetic wave is to be transmitted to a satellite or vice versa. It has been found that the optimum frequency for interplanetary links lies between 100 and 300 Mc/s. In this frequency range, the noise factor of receivers is of

the order of 2 or 3, with the result that the corresponding noise power is equal to

$$P_N = 1{\cdot}2 \times 10^{-20} \text{ watts/cycle},$$

and the cosmic noise power can be written

$$P_{\cos} = A \times 10^{-22} \text{ watts/cycle},$$

A is the surface equivalent to the paraboloid of revolution of the terrestrial aerial; $A = S/2$. The minimum total noise is achieved by arranging that

$$P_N = P_{\cos}$$

which implies a surface area of 240 m².

IV.1.3. *Pulse Noise*

Pulse noise is caused by external phenomena such as electrical discharges, static atmospheric electricity and atmospheric phenomena.

The analytic expression for a pulse can be expanded in the form of a Fourier series, but only a fraction of the components forming the series can enter the receiver. Everything depends upon the bandwidth of the latter and the amplitude of the useful signal.

When the repetition frequency of the pulses is constant and equal to F, the bandwidth of the receiver is liable to accept any harmonics between nF and $(n + k)F$. The spectrum can be regarded as continuous only if the pulses occur at a far lower rate than the central frequency of the receiver, and one is led to study the effect of white noise on the behaviour of the useful signal. If this repetition frequency is greater, the problem becomes a more complicated one, and it can be shown that the signal-to-noise ratio passes through a minimum, the value of which is of the form

$$\left(\frac{S}{N}\right)_{\min} = \frac{1}{\Delta\omega} \cdot \frac{E}{\sqrt{2}} KQ\, e^{\frac{1}{(KQ)^2}} \qquad \text{(IV.7)}$$

in which K is the coupling coefficient between the receiver circuits, E is the amplitude of the useful signal, and $2\Delta\omega$ is the bandwidth of the receiver.

When the repetition frequency of the pulses is random, the calculations rapidly become unmanageable.

It should be emphasized however that in these cases protection against such noise can be efficiently ensured by reducing the receiver bandwidth to a minimum.

IV.1.4. *Calculation of the Signal-to-Noise Ratio for Various Different Types of Modulation*

In this it is assumed that the noise in question is a fluctuating noise, the high frequency part of whose spectrum is uniform; the sensitivity of the receiver to noise will be characterized by a coefficient K.

In the case of ordinary amplitude modulation, the signal-to-noise ratio takes the form

$$\left(\frac{S}{N}\right)_{\text{AM}} = \frac{m^2}{4} \quad \frac{\text{peak power}}{2K\mathscr{B}} \tag{IV.8}$$

in which $\mathscr{B}$ is the bandwidth occupied by the modulation and m is the modulation index.

To compare the various signal-to-noise ratios with that of amplitude modulation, we must define a typical modulation index which we shall call 100 per cent, and (IV.8) becomes

$$\left(\frac{S}{N}\right)_{\text{AM}} = \frac{\text{carrier power}}{2K\mathscr{B}}.$$

The power of the carrier wave, P_p, is in fact equal to

$$P_p = \frac{m^2}{4} \times P_{\text{peak}}.$$

In this expression, we have assumed implicitly that there is no correlation between the noise components in the two sidebands. If this were not so, there would be parasitic modulation at the transmitter, and the signal-to-noise ratio would be less than the value given by the equation.

IV.1.4.1. *The Signal-to-Noise Ratio in SSB Modulation*

The ratio takes the form

$$\left(\frac{S}{N}\right)_{\text{SSB}} = \frac{P_{\text{peak}}}{K\mathscr{B}}.$$

The signal-to-noise ratio efficiency is defined to be

$$R_{\text{SSB}} = \frac{\left(\frac{S}{N}\right)_{\text{SSB}}}{\left(\frac{S}{N}\right)_{\text{AM}}}$$

or

$$R_{SSB} = \frac{\dfrac{P_{peak}}{K\mathscr{B}}}{\dfrac{\frac{1}{8}P_{peak}}{K\mathscr{B}}} = 8.$$

The gain of single sideband modulation over ordinary amplitude modulation is therefore

$$R_{SSB} = 8 = 9 \text{ db}.$$

IV.1.4.2. *The Signal-to-Noise Ratio in FM Modulation*

Let f be the frequency of the carrier wave; a noise component of frequency f_b will enter the receiver if it lies within the bandwidth of the latter. The signal-to-noise ratio takes the value

$$\left(\frac{S}{N}\right)_{FM} = \frac{P}{2KF_{max}} \times 3\,\frac{\Delta f^2}{(F_{max})^2}$$

in which Δf is the variation of the carrier frequency and F_{max} is the maximum frequency of the modulation. The gain with frequency modulation is therefore

$$R_{FM} = 3\,\frac{\Delta f^2}{(F_{max})^2}. \qquad \text{(IV.10)}$$

The ratio $\Delta f/F_{max}$ represents the modulation index, and the improvement is more pronounced as this coefficient is increased.

IV.1.4.3. *The Signal-to-Noise Ratio in Pulsed Modulation (PAM)*[53]

If a signal plus noise is sampled at a repetition frequency f_r by means of short pulses, we can show that after sampling and filtering (with a low-pass filter of cut-off frequency $f_r/2$) the signal-to-noise ratio is equal to its value before sampling.

It should be noted that sampling a signal on which noise is superimposed reduces the signal-to-noise ratio by a factor F_{max}/δ in which F_{max} is the maximum LF frequency to be transmitted and δ is the limit of the noise band $(0 - \delta)$.

PAM modulation. When the noise is white, with constant spectral density $b(f) = b$, the signal-to-noise ratio for PAM modulation is

$$\left(\frac{S}{N}\right)_{PAM} = \frac{f_r}{2F_{max}}\,\frac{P}{b(f)\mathscr{C}'}$$

in which P is the mean power in the samples,

f_r the repetition frequency of the sampling,

$f_c = \mathscr{B}'$, the useful bandwidth of the pulses $(= \frac{1}{2}\theta)$.

On introducing the duty factor $\theta/2$, the efficiency, in comparison with amplitude modulation, is equal to

$$R_{\text{PAM}} = \frac{\theta}{T}.$$

This is true only provided the carrier powers during the transmission of the pulses and during the amplitude modulation are identical.

PAM–FM modulation. When the frequency of the carrier wave is modulated, one benefits from the advantages offered by this type of modulation, and the overall efficiency is equal to the product of the efficiencies, thus:

$$\begin{aligned} R_{\text{PAM–FM}} &= R_{\text{PAM}} \times R_{\text{FM}} \\ &= \frac{\theta}{T} \times 3\frac{\Delta F^2}{f_c^2}. \end{aligned} \tag{IV.11}$$

Alternatively, this can be written

$$\left(\frac{S}{N}\right)_{\text{PAM–FM}} = \left(\frac{S}{N}\right)_{\text{PAM}} \times \frac{3\Delta F^2}{f_c^2}. \tag{IV.12}$$

IV.1.4.4. *Signal-to-Noise Ratio in PPM and PDM Modulation*

The noise appears as a lack of precision in the positioning of the leading or falling edge of the pulse. The noise peaks must occur with a probability of less than 1/1000 if the perturbation which they produce is to be negligible. We must introduce the requirement that the power of the carrier during the pulse exceeds twenty times the noise power during the same period.

If ε denotes the rise-time of the pulse and ΔT the maximum displacement due to the modulation, the signal-to-noise ratio takes the form:

$$\left(\frac{S}{N}\right)_{\text{PPM}} = \frac{P_{\text{peak}}}{P_{\text{noise}}}\left(\frac{\Delta T_{\text{effective}}}{\varepsilon}\right)^2.$$

For sinusoidal modulation,

$$\Delta T_{\text{effective}} = \frac{\Delta T}{\sqrt{2}}$$

and

$$\left(\frac{S}{N}\right)_{\text{PPM}} = \frac{P_{\text{carrier}}}{P_{\text{noise}}}\left(\frac{\Delta T}{\varepsilon}\right)^2.$$

The gain with this kind of modulation is thus:

$$R_{\text{PPM}} = R_{\text{PAM}} \times \left(\frac{\Delta T}{\varepsilon}\right)^2 = \frac{\theta}{T}\left(\frac{\Delta T}{\varepsilon}\right)^2. \tag{IV.13}$$

IV.1.4.5. *The Signal-to-Noise Ratio in Coded Modulation (PCM)*[53]

At this stage, quantization noise, also termed granularity noise, will be left aside and considered later with distortion noise. In the present section we only attempt to define the signal-to-noise ratios which result from fluctuating noise, and not those which are dictated by the modulation itself.

We denote by A the amplitude of the pulses after detection. The pulses are "read" at the level $A/2$; a pulse, equal to or greater than $A/2$ in amplitude, indicates the presence of the digit "1", while the contrary indicates the digit "0". Noise, greater in amplitude than $A/2$, will thus clearly introduce errors.

The probability distribution of the noise amplitude x is

$$P(x) = \frac{1}{\Gamma\sqrt{2\pi}} e^{-\frac{x^2}{2\Gamma^2}},$$

in which Γ is the root mean square value of x:

$$\Gamma^2 = \int_{-\infty}^{\infty} x^2 P(x)\, dx.$$

The probability of the level being greater than $A/2$ is therefore

$$P\left(\frac{A}{2}\right) = \int_{A/2}^{\infty} P(x)\, dx$$

or finally,

$$P\left(\frac{A}{2}\right) = 1 - \Pi\left(\frac{A}{2\Gamma}\right)$$

in which $\Pi(x)$ is the "error function". If $V, V/2, \ldots, V/2^n$ are the "weights" associated with the different pulses in the code, the quantization levels of the signal vary between 0 and $2V$. The noise creates false pulses which occur with the probability $P(A/2)$.

It can be shown that the mean noise power caused by spurious pulses is $P(A/2)\ V^2$ and for the series of pulses with different weights:

$$P_N \simeq P\left(\frac{A}{2}\right)\left[V^2 + \frac{V^2}{2^2} + \cdots + \frac{V^2}{2^{2n}}\right] \simeq \frac{4}{3} P\left(\frac{A}{2}\right) V^2.$$

The signal-to-noise ratio is therefore

$$\left(\frac{S}{N}\right)_{PCM} = \frac{\frac{V^2}{2}}{\frac{4}{3} P\left(\frac{A}{2}\right) V^2} = \frac{3}{8} \frac{1}{P\left(\frac{A}{2}\right)} = \frac{3}{8\left[1 - \Pi\left(\frac{A}{2\Gamma}\right)\right]}. \quad \text{(IV.14)}$$

The value of $A/2\Gamma$ is proportional to the square root of the signal-to-noise ratio at the input of the demodulator:

$$\left(\frac{S}{N}\right)_{\text{input}} = \frac{2A^2}{\Gamma^2}.$$

In Fig. IV.5, curves are plotted which relate the values of the signal-to-noise ratio at the input and output; it is clear that the improvement increases very rapidly beyond the 12 db value of the signal-to-noise ratio.

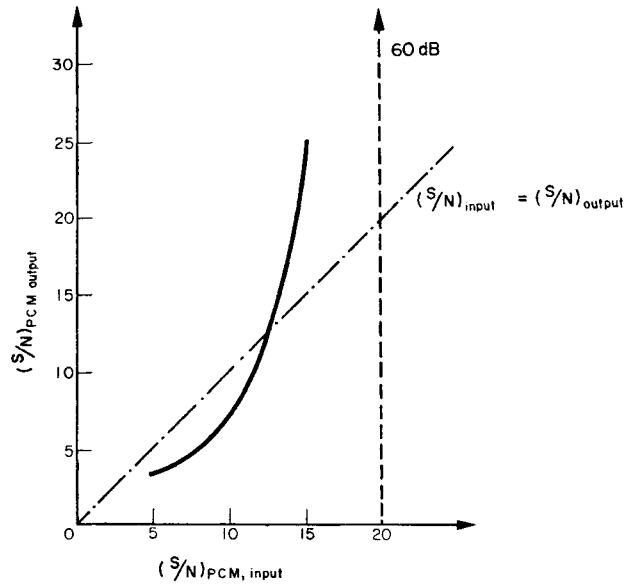


FIG. IV.5. Signal-to-noise ratio at the output as a function of the signal-to-noise ratio at the input.

IV.2. INTERFERENCE AND JAMMING PHENOMENA

IV.2.1. *General Introduction*

The two curves which correspond to the selectivity of the circuit and to the spectrum of a jamming signal are plotted in Fig. IV.6. The region AB, covering the bandwidth $\mathscr{B}$, in which the useful modulation is to be found, is directly affected by every component of the spectrum of the jamming signal. The possibility of jamming depends upon the selective properties of the circuit.

Jamming procedures will be dealt with in the following paragraphs. The phenomenon of jamming can itself always be split into one of interference between two waves, which is the way in which the problem is approached here.

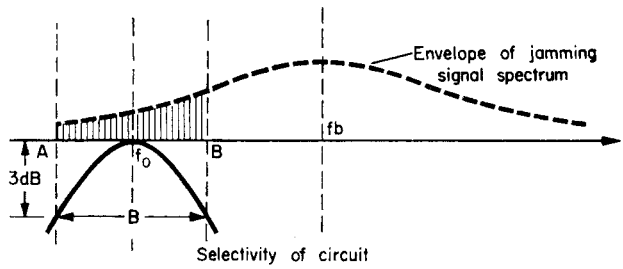


FIG. IV.6. Effect of a jammer on a selective circuit.

IV.2.2. *Interference Phenomena*

Suppose that

$$e_1 = E_1 \sin \omega_1 t$$

$$e_2 = E_2 \sin \omega_2 t$$

are two components which combine to form an electric signal; their resultant can be written in the form:

$$e = E_1[\sin \omega_1 t + x \sin \omega_2 t]$$

with $x = E_2/E_1$. The instantaneous frequency, f is given by

$$f_i = f_1 + f_b \left[\frac{x^2 + x \cos (\omega_1 - \omega_2)t}{1 + x^2 + 2x \cos (\omega_1 - \omega_2)t} \right], \qquad \text{(IV.15)}$$

in which $f_b = f_2 - f_1$ is the beat frequency. Expanding each of the terms between brackets in series form gives

$$f_i = f_1 + f_b \sum_1^\infty (-1)^{n+1} x^n \cos n(\omega_1 - \omega_2)t,$$

$$f_i = f_2 - f_b \sum_1^\infty (-1)^{n+1} x^n \cos n(\omega_1 - \omega_2)t.$$

The beat frequency and its harmonics lie either outside the frequency band of the receiver, or within it. In the first case, the signal with the greater ampli-

tude predominates. In the second case, the weaker signal makes a contribution to the noise, and its effective value is

$$N_{\text{eff}} = \frac{f_b}{\sqrt{2}} \sqrt{\sum_{n=1}^{n_0} b_n^2};$$

n_o represents the order of the harmonic which is capable of being accepted by the receiver band, and $n_0 f_b$ is less than or equal to the cut-off frequency of the receiver.

In the case of FM modulation, for example,

$$\left.\begin{aligned} (f_i)_{\min} &= f_1 - \frac{x f_b}{1 - x} \\ \text{and} \qquad (f_i)_{\max} &= f_2 + \frac{x f_b}{1 - x}. \end{aligned}\right\} \qquad \text{(IV.16)}$$

The total frequency deviation becomes

$$(f_i)_{\max} - (f_i)_{\min} = f_b \frac{1 + x}{1 - x},$$

and this is infinite for $x = 1$, resulting in a very marked distortion at the discriminator stage.

Disturbance due to interference thus becomes more pronounced as the beat frequency f_b increases, and as the ratio x tends towards 1: when, that is to say, the amplitude of the perturbing signal becomes equal to that of the useful signal.

Consider now the case in which two waves which originated in the same source but which have followed different paths are re-combined. At a given point in space, the two electric field intensities

$$e_1 = E_1 \sin \omega t \quad \text{and} \quad e_2 = x E_1 \sin \omega(t - \Delta t)$$

combine to form the resultant field, e_R, which takes the form

$$e_R = E_1 \sin \omega t \,(1 + x \cos \omega \Delta t - x \sin \omega \Delta t \cot \omega t).$$

When $\omega \Delta t = \Delta \varphi = k\pi$, we have

$$e_R = E_1 \sin \omega t [1 \pm x],$$

and the relative error x affects only the amplitude of the original wave. The value of x is usually small enough for its influence to be negligible, in this case. When, on the contrary $\omega \Delta t = \Delta \varphi = k\pi/2$, we find

$$e_R = E_1 \sin \omega t [1 \pm x \cot \omega t]$$

so that the signal is heavily distorted.

This interference phenomenon can easily be observed in experiments carried out over the sea. If the field intensity at a receiver aboard an aeroplane is recorded it is found to oscillate between a maximum and a minimum as shown in Fig. IV.7. The sea provides an ideal reflecting surface for electromagnetic radiation.

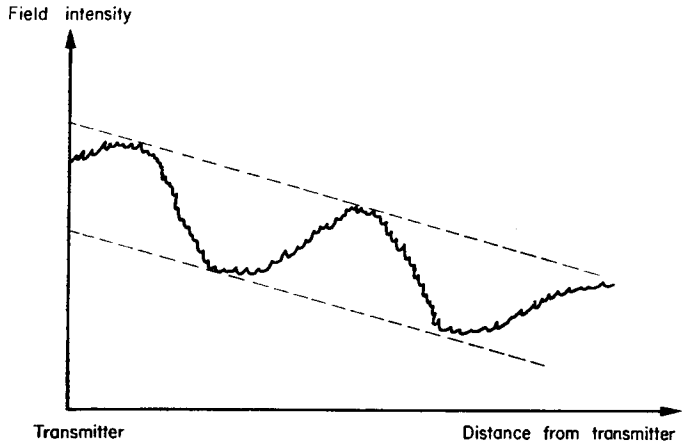


FIG. IV.7. Example of RF field intensity received over the sea as a function of distance.

IV.2.3. *Jamming*

The jamming of any instrument may be the result of a parasitic emission at a frequency close to that of the useful frequency—this is simply the case of two-wave interference, however.

When the jamming is deliberate, it usually consists of an emission "wobulated" across a fairly wide frequency range as Fig. IV.8 shows. The jamming signal is emitted at a frequency which varies between two values, f_1 and f_2 say. The frequency range $|f_2 - f_1|$ is scanned for a time T. The central transmitter frequency and the receiver bandwidth are denoted by f_0 and $\mathscr{B}$ respectively. The disturbance jams the useful signal for a time $2\Delta T$, at intervals of time $1/T$. The narrower the bandwidth the shorter is the time $2\Delta T$ during which the disturbing influence acts. When the slope of the jammer, defined by $(f_2 - f_1)/T$, is slight, the time taken to cross the band is long; on the other hand, the repetition frequency is less, and the receiver is jammed at a slower rate. In any case, we always use as narrow a bandwidth as possible, compatible with the modulation employed and the time constant necessary for good transmission of the data.

Consider now a body in motion along an arbitrary trajectory between two

points A and B. We suppose that at B, there is a jammer (see Fig. IV.9), and that the missile is piloted from some other point, C.

The powers emitted at C and B are P_C and P_B respectively; the receiver at A receives the two waves P_{RC} and P_{RB} for which

$$\frac{P_C}{P_{RC}} = \frac{4\pi d_1^2}{G_C S_{RC}} \quad \text{and} \quad \frac{P_B}{P_{RB}} = \frac{4\pi d_2^2}{G_B S_{RB}}$$

d_1 and d_2 are the distances AC and AB respectively, G_C and G_B are the gains of the transmitter aerials at C and B respectively, which are assumed equal.

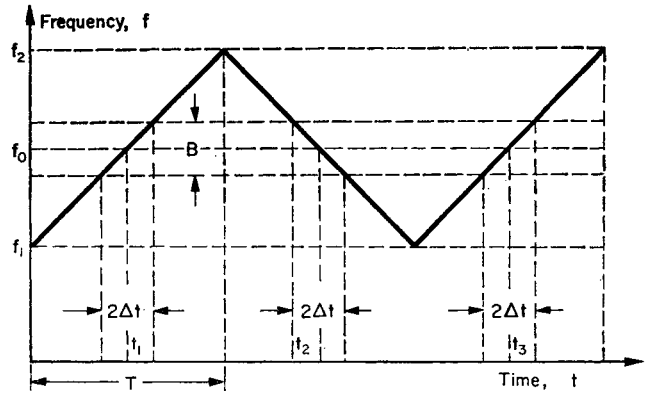


FIG. IV.8. Wobulation of a jammer.

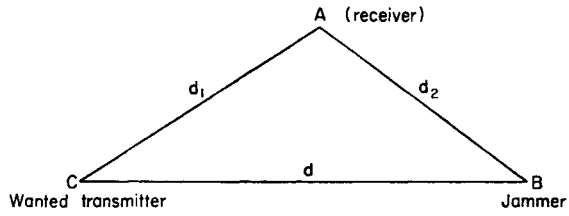


FIG. IV.9. The positions of the moving point A with respect to the jammer B and the wanted transmitter C.

S_{RC} and S_{RB} are the effective areas of the receiver aerial with respect to the transmissions from C and from B respectively, which are taken to be equal. Thus,

$$\frac{P_C}{P_B} \cdot \frac{P_{RB}}{P_{RC}} = \left(\frac{d_1}{d_2}\right)^2.$$

The ratio $P_C/P_B = p$ is constant; the jamming is effective only when the ratio P_{RB}/P_{RC} is greater than or equal to a certain value termed the jamming coefficient, k_{limit}. The jamming limit is therefore defined by

$$\left(\frac{d_1}{d_2}\right)^2 = k_{\text{limit}} \times p = m^2$$

or $(d_1/d_2) = m$ where m is a constant ($m = \sqrt{pk_{\text{limit}}}$). The minimum distance at which jamming begins is such that

$$\left(\frac{d_2}{d_1}\right)_{\text{min}} = m, \quad \frac{(d_2)_{\text{min}}}{d - (d_2)_{\text{min}}} = m,$$

$$(d_2)_{\text{min}}(1 + m) = md,$$

$$(d_2)_{\text{min}} = \frac{m}{1 + m} d.$$

IV.3. ERRORS AND THRESHOLDS OF MEASUREMENT

These errors can be classified into two different sorts: static and dynamic. The latter are related to the variations of the input magnitude: they are a function of the applied signal.

The notion of static error is a very restricted one; it really only comprises calibration inaccuracy. Nevertheless, very slow drifts—quasi-static effects—can be regarded as eligible for this class of error.

IV.3.1. *General Points*

Suppose that a state is made to correspond closely to the exact value of a steady and measurable physical quantity; and that this state is characterized by a number, which is a measure of the physical quantity. It is not possible to make such a measurement unless the speed at which the physical quantity G varies, as a function of the number N which represents measurement of the quantity, has a finite value. Suppose that this speed is a constant, so that

$$\frac{\mathrm{d}G}{\mathrm{d}N} = k;$$

on integrating, we obtain

$$N = \frac{1}{k}(G - G_R) + N_R,$$

in which N_R is a specific number, corresponding to the reference value G_R of the physical quantity. The measurement is always in the form of a difference between the quantity to be measured and a reference quantity.

The relative error in measuring N is thus

$$\left|\frac{\Delta N}{N}\right| = \left|\frac{\Delta(G - G_R)}{G - G_R}\right| + \left|\frac{\Delta N_R}{N_R}\right|.$$

This is composed of the relative errors in the physical quantity to be measured and in the reference. A telemetry chain consists of a measuring device, a translator, an amplifier and a modulator together with other elements as shown schematically in Fig. IV.10.

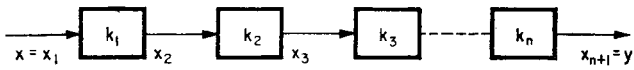


FIG. IV.10. Telemetry chain.

The input magnitude x and the output magnitude y are related by the expression

$$y = k_1 \cdot k_2 \ldots k_n x$$

or

$$\Delta x = \frac{\Delta y}{k_1 \cdot k_2 \ldots k_n}.$$

The indeterminacy of a physical quantity is defined as the smallest value of it which can be measured.

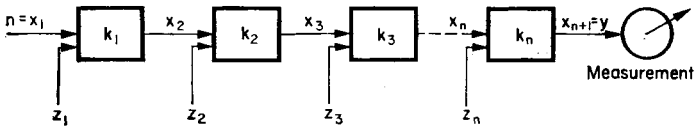


FIG. IV.11. Measurement chain with references.

The reference quantities can be introduced into the measurement chain, in the way shown in Fig. IV.11. Taking the uncertainty unit as the error in reading the measuring instrument, the effect of this error upon the input magnitude, x, is found to be

$$\Delta x = \Delta x_1 = \frac{\Delta x_2}{k_1} + \Delta z_1$$

$$\Delta x_2 = \frac{\Delta x_3}{k_2} + \Delta z_2$$

$$\Delta x_{n-1} = \frac{\Delta x_n}{k_{n-1}} + \Delta z_{n-1}$$

$$\Delta x_n = \frac{\Delta x_{n+1}}{k_n} + \Delta z_n,$$

or

$$\Delta x = \frac{\Delta y}{k_1 \cdot k_2 \dots k_n} + \frac{\Delta z_n}{k_1 \cdot k_2 \dots k_{n-1}}$$

$$+ \frac{\Delta z_{n-1}}{k_1 \cdot k_2 \dots k_{n-2}} + \dots + \Delta z_1.$$

Defining the quantity

$$\varrho = \frac{\Delta x}{\Delta y/(k_1 \cdot k_2 \dots k_n)}$$

to be an "error index" (referring the measurements back to the input to the system) we have

$$\varrho = 1 + k_n \Delta z_n + k_n k_{n-1} \Delta z_{n-1} + \dots + (k_n k_{n-1} \dots k_1) \Delta z_1$$

An error index as nearly as possible equal to unity is aimed at, in order to give the whole system its best configuration. It should be emphasized that to attain a value of ϱ close to unity requires stabilization of the reference quantities.

Throughout this discussion the possibilities of saturation and of non-nearity in the transfer characteristics of the measurement chain are excluded.

IV.3.2. *Static Errors*

This type of error is inherent in the discontinuous features encountered in every instrument, a few examples of which are given below.

The thickness of the trace of a light beam which activates a photographic plate is finite, and this is a restriction, therefore, on the interpretation of an image. The hair-line of a measuring instrument, the interval between two definite marks, the breadth of the needle of a galvanometer, all lead to inaccuracy in measurement. The stability of an oscillator is limited to a certain value. If a quartz oscillator, accurate to one part in 10^4, is employed in an oscillator circuit, the accuracy of the frequency f provided will be one part in 10^4, which means that the error so far as frequency is concerned is f parts in 10^4. In electronic circuitry, such elements as resistors and capacitors are employed, and the values of these are usually only accurate to within 5 or 10 per cent.

Calibration errors are very serious, since they have repercussions along the whole measuring chain.

It has already been seen that the error due to quantization is less than or equal to the distance between two levels of quantization. We shall now examine the interpolation error introduced when a true curve is represented by a succession of horizontal steps. See Fig. IV.12.

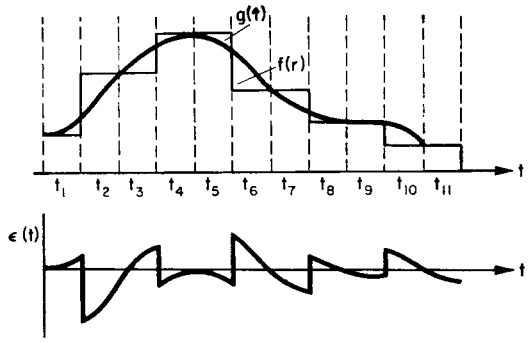


FIG. IV.12. Interpolation error.

Let $g(t)$ be the real continuous function, for which the discontinuous function $f(t)$ is substituted. At any instant, the error introduced is

$$\varepsilon(t) = g(t) - f(t).$$

If Δt is the sampling interval: $t_1 = \Delta t$, $t_2 = 2\Delta t$, ... $t_n = n\Delta t$, ..., and

$$\varepsilon_1(t) = g(t) - g(0) \quad \text{for} \quad 0 < t < t_1$$

$$\varepsilon_3(t) = g(t) - g(t_2) \quad \text{for} \quad t_1 < t < t_3$$

$$\varepsilon_5(t) = g(t) - g(t_4) \quad \text{for} \quad t_3 < t < t_5.$$

The root mean square error, therefore, is given by

$$\begin{aligned}
\overline{\varepsilon^2} &= \int_0^{t_1} \varepsilon_1^2(t)\,dt + \int_{t_1}^{t_3} \varepsilon_3^2(t)\,dt + \cdots \\
&= \int_0^{t_1} [g(t) - g(0)]^2\,dt + \int_{t_1}^{t_3} [g(t) - g(t_2)]^2 dt + \cdots \\
&= \int_0^{\infty} [g(t)]^2\,dt \\
&\quad + 2\Delta t \left\{ \frac{g^2(0)}{2} + g^2(2\Delta t) + g^2(4\Delta t) + \cdots + g^2[(2n-1)\Delta t] \right\} \\
&\quad - 2g(0) \int_0^{\Delta t = t_1} g(t)\,dt - 2g(2\Delta t) \int_{t_1}^{t_3} g(t)\,dt \\
&\quad - \cdots .
\end{aligned}$$

The mean error is

$$\varepsilon_{\text{mean}} = \frac{1}{n\Delta t}\left[\int_0^{n\Delta t} [g(t) - f(t)]\,\mathrm{d}t\right]_{n\Delta t \to \infty}$$

$$= [g(t)]_{\text{mean}} - \frac{1}{n}[g(0) + 2g(\Delta t) + 2g(2\Delta t) + \cdots].$$

It is sometimes easy to calculate $g(t)$ if this is a simple function, but usually it fluctuates in a random manner; it is then necessary to proceed graphically.

IV.3.3. *Dynamic Errors*

The phenomena of vibrations, shocks, acceleration and temperature, amongst others, all have a particularly important influence on the behaviour of the measuring instruments and electronic instruments which are used in both remote control and telemetry. These perturbing influences produce errors not only in the initial calibration but also during the actual measurement itself.

Vibrations, shocks and acceleration all belong to the same class of phenomena, which are related directly to the accelerating force.

The frequency spectrum corresponding to a periodic vibration contains a series of discrete components for each vibration frequency.

Muffled noise, for example, originating from an ill-defined source, yields a uniform spectrum.

We shall regard shocks as sudden non-periodic changes of acceleration, which allows their effect on the behaviour of remote-control and telemetry systems to be analysed.

IV.3.3.1. *The Effect of Acceleration on the Behaviour of Radio Components*

Two elements, connected together elastically, which have different masses will become displaced with respect to one another when the combined system is subjected to an acceleration. The grid-cathode space of the radio valves would be modified, for example, and the electron density emerging from the filament would be modulated. The result would appear as a fluctuation noise, more or less periodic according to the system inquestion, and this would be superimposed onto the noise of the electronic devices themselves. Further, the grid-cathode capacity of such a valve would also undergo this modulation, and the circuit connected across its terminals would reflect the modulation in synchronism.

The movement of the grid relative to the cathode is betrayed by changes in the working point of the valve, in the amplification and in the interelectrode capacitances. This last effect is the most serious; in telemetry, for example,

it may introduce parasitic frequency modulation which can be very troublesome.

Acceleration effects can, however, be compensated by orientating the vacuum tubes in a suitable manner. From this point of view, transistors are undoubtedly preferable.

The problem of shocks and of sudden changes in acceleration is less important where radio equipment is concerned. These shocks only occur, for the most part, when the instrument is not in use.

Transistors, on the contrary, can bear very sharp mechanical shocks and jars without any harmful result, since their weight is very slight.

When the missile in which the electronic equipment is mounted is fired, the elements of the circuits all undergo the same constant acceleration. This appears as an error in any circuits which are sensitive to direct currents. Suppose, for example, that the diaphragm of a barometer is linked to the wiper of a potentiometer; the inertial force will thus create a measuring error. It is thus preferable to place the axis of the diaphragm perpendicular to the direction of the accelerating force.

We have now seen what kinds of effect movements due to acceleration (shocks, vibrations, steady acceleration, noise for example) may have on components on board a missile.

We shall now consider the problem theoretically.

IV.3.3.2. *Characteristics of the Error produced by Acceleration*

In order to isolate an object from the perturbations produced by shocks and vibrations, the function relating the frequency response and the relative displacement must be modified.

This can be done by means of an elastic support.

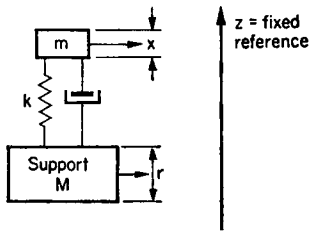


FIG. IV.13. Elastically coupled system.

Let m be the mass to be isolated, which is connected to a support by a substance which has a coefficient of elasticity k and a damping coefficient C. The whole system is represented schematically in Fig. IV.13.

Consider the effect of applying vibrations $r = R \sin \omega t$ to the mass M of the support; if $m \ll M$, the resonance frequency of the mass m and the damping coefficient, ω_c and a respectively, are given by

$$\omega_c = \sqrt{\frac{k}{m}} \tag{IV.17}$$

$$a = \frac{C}{2\sqrt{km}}. \tag{IV.18}$$

For critical damping, the coefficient a is equal to unity.

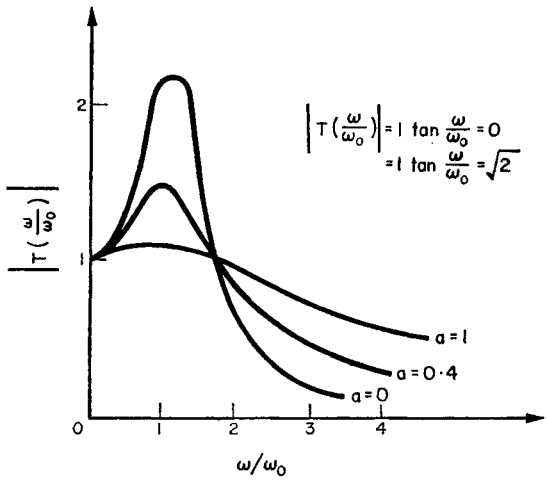


FIG. IV.14. Frequency response for various values of damping coefficient a.

The amplitude of the displacement of the mass m, corresponding to a displacement R of the mass M, is taken as X, and the ratio of the amplitudes of the transmitted force to the applied force can be defined as a force transmission coefficient. This quantity can be identified with the amplitude transfer function,

$$\left|T\left(\frac{\omega}{\omega_0}\right)\right| = \frac{X(\omega)}{R(\omega)} \tag{IV.19}$$

shown plotted in Fig. IV.14.

If γ is the value of the acceleration applied to the system, and g_γ is the component of gravity in the direction of γ, the displacement, δ, of m is

$$\delta = m(\gamma + g_\gamma)/k = \frac{\gamma + g_\gamma}{\omega_0^2}. \tag{IV.20}$$

As an example of this, consider the motion of the armature of a relay, whose return spring coefficient is k and whose damping to all intents and purposes is zero. The transfer of vibration to the armature will be negligible if $\omega/\omega_0 \gg 1$.

In the case of the grid-cathode space of a radio valve, the mass m of the grid can be regarded as linked elastically to the mass μ of the cathode, and the latter is in turn connected mechanically to the support M. The displacements of m, μ and M are denoted by x, y and r respectively. If $F(\omega)$ is the variation of the support position with frequency, the overall response is

$$|T_x| \cdot |T_y| \cdot F(\omega),$$

and

$$\omega_{0x} = \sqrt{\frac{k_x}{m}}$$

$$\omega_{0y} = \sqrt{\frac{k_y}{\mu}}$$

where $m \ll \mu$, $\omega_{0x} \gg \omega_{0y}$.

To avoid relative motion between grid and cathode, the condition $(x - y) = 0$ must be satisfied irrespective of the value of the applied frequency, ω.

In fact, the condition that we try to satisfy is

$$|T_x - 1| \cdot |T_y| \ll 1. \tag{IV.21}$$

In practice, a value

$$|T_x - 1| \cdot |T_y| < 3 \times 10^{-2}$$

is accepted, for $\omega < 8\omega_{oy}$.

It must also be noted that the function $F(\omega)$ depends upon the type of acceleration which is applied. In the case of vibrations, $F(\omega)$ is sinusoidal; for noise, $F(\omega)$ represents a constant spectral density. When sudden shocks are applied $F(\omega)$ has the spectrum of a pulse.

When the rockets which drive the missile are fired, the contents undergo a sudden burst of acceleration which corresponds to a unit step.

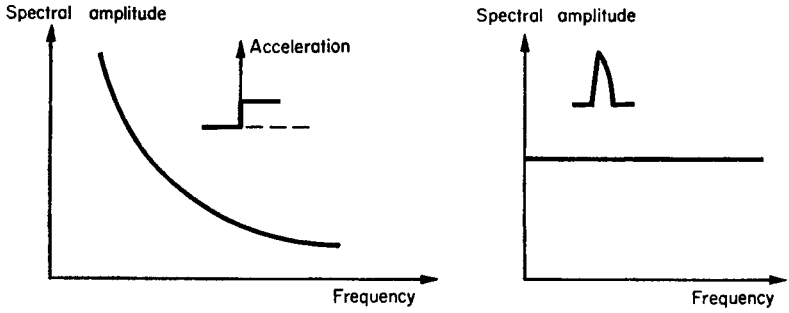


FIG. IV.15. Frequency spectrum of a unit step and a Dirac δ-function.

The various types of spectrum are shown in Fig. IV.15. As can be seen, a pulse is rich in frequency components, and this necessitates a higher value of the ratio ω_{0x}/ω_{0y}.

IV.3.3.3. *Errors due to Temperature Effects*

In every list of specifications for components which are to be flown in an aeroplane or a missile, a special clause about the effect of temperature is always inserted.

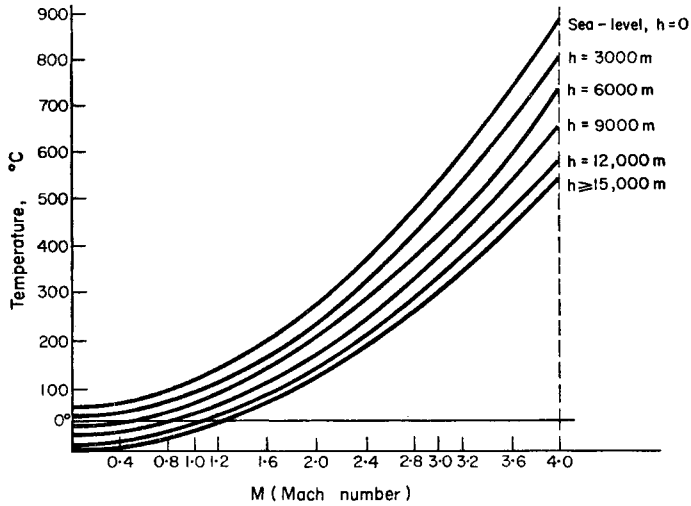


FIG. IV.16. Increase of skin temperature of a missile, as a function the Mach number, for various altitudes.

The theoretical value of the temperature of the hull-skin of a missile, travelling through the atmosphere at a speed of about Mach 2 at an altitude of 12,000 m, may reach 150 °C. What is more, the same missile may have been launched from a base at either the equator or one of the poles.

In Fig. IV.16 curves are plotted showing the hull-skin temperature of the missile as a function of the speed and altitude.

In fact, the temperature inside the missile is more stable, since the thermal capacity of the total equipment is such that the temperature inside the missile will not reach the temperature of the skin, when the latter is 150 °C, until an interval of time considerably longer than the time of flight of the actual missile has elapsed.

Valve heaters, the rotation of the motors and dissipation in resistors, all cause appreciable heating of the equipment.

Generally speaking, the radio equipment should be able to support a temperature range of − 50 °C to + 100 °C without being adversely affected. To

satisfy these exacting conditions every possible advantage must be taken of advances in the technology of radio equipment components. For example, germanium transistors or diodes can only support junction temperatures up to 85°, whereas silicon transistors can work at temperatures up to 150°.

Temperature errors can be classified into three kinds: calibration or reference errors, deterioration of the noise factor, and performance errors. Consider the capsule of a barometer, indicating a static pressure (corresponding to the altitude of the missile) by means of telemetry. The air in the capsule expands under the effect of a rise in temperature, and the altitude zero (sea-level) will have to be re-adjusted according to the temperature at any moment. Further, even after calibration, when the missile is launched, the temperature of the medium surrounding the capsule rises, and the altitude recorded is affected by the pressure change Δp caused by the heating ΔT °C which occurs during the course of the flight of the missile. It is then necessary to compensate for the temperature variations by thermostatic temperature control. This is not a complete answer since the capsule receives ambient air of a rapidly varying temperature through a Pitot tube.

Suppose that a continuous current variation is to be transmitted. If a varicap (a varicap has a capacity which varies with the current in the semiconductor) is used as the transducer, it is possible to modulate the frequency of a signal. When the temperature increases, the quiescent current i_0 of the diode also increases in an exponential manner, and this is added to the useful current of the input signal I_e.

If $C = ki$ is the variation law of the varicap, the error in the capacity, and hence in the frequency, is

$$\Delta C = \Delta(ki_0) = k' \, e^{\mu T} \cdot \Delta T$$

which is a function of temperature.

The capacity variation which is produced through the effect of a temperature change can be compensated by connecting the current transducer in a thermally balanced bridge, or alternatively, by selecting components which have a temperature coefficient of opposite sign and placing these in series with the varicap in question. The effect of temperature also makes itself felt as a drift in gain characteristics and as frequency shifts in tuned circuits. A tuned circuit comprises either transmission lines for UHF circuitry or inductors in parallel with capacitors. All these elements expand to a greater or lesser extent, and this alters their values by an amount which varies as a function of the temperature; they thus become out of tune.

The permeability of magnetic media (ferrites, for example) is also a function of temperature. On approaching relatively high temperatures, severe changes in gain often result. This can be characterized by the Curie point, which for some ferrites lies near 90 °C.

With transistors, this problem of temperature is a critical one. Their quiescent current is, as already mentioned, exponentially dependent upon

the temperature. This effect can be partially compensated for by means of a judicious use of feedback or by using thermistors; inevitably, however, a transistor (or more generally, a semi-conductor) which accepts a maximum dissipation power of P watts at 25 °C can only dissipate on average $P/2$ watts at 50 °C.

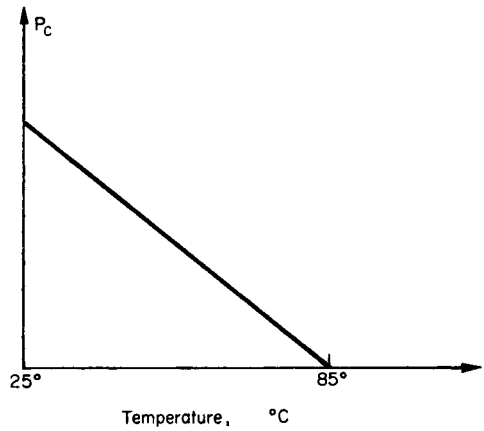


FIG. IV.17. Variation of the power dissipated at the collector, P_c, as a function of the ambient temperature.

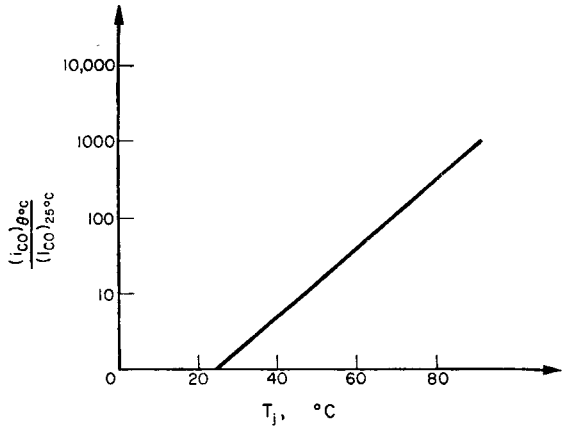


FIG. IV.18. Variation of the ratio of quiescent current at 0°C to quiescent current at 25°C as a function of the junction temperature T_j°C.

In Fig. IV.17 we show the dissipation curve as a function of temperature, and in Fig. IV.18 the relative changes in the quiescient current as a function of the junction temperature of the transistor.

The third effect of temperature is to increase the level of fluctuation noise, but this phenomenon has already been dealt with in the section on noise.

IV.4. DISTORTION DUE TO NON-LINEARITY

In general, no transmission system is perfectly linear. If a sinusoidal signal of the form $e^{j\omega t}$ is applied to the input of a quadripole, the transfer function of which is $K(j\omega)$, a signal $K(j\omega)\, e^{j\omega t}$ results at the output. The transfer function $K(j\omega)$ can be written in the form

$$K(j\omega) = A(\omega)\, e^{j\psi(\omega)}. \tag{IV.22}$$

A broad division of the distortion due to non-linearity can be made into those which affect the amplitude characteristics and those which affect the phase characteristics.

IV.4.1. *Amplitude Distortion*

The relation between a signal S at the output and the signal E at the input can always be written in a power series form as

$$S = a_1 E + a_2 E^2 + a_3 E^3 + \cdots + a_n E^n. \tag{IV.23}$$

If the transmission is not perfectly linear, the coefficients $a_2, a_3, \ldots, a_n$ are not zero, and their values depend upon the nature of the non-linearity.

If the signal E is a circular function

$$E = e^{j\omega t}$$

then substituting this value of E into equation (IV.23) yields

$$\begin{aligned} S &= a_1 e^{j\omega t} + a_2 e^{j2\omega t} + \cdots + a_n e^{nj\omega t} \\ &= a_1 e^{j\omega t}\left[1 + \frac{a_2}{a_1} e^{j\omega t} + \frac{a_3}{a_1} e^{2j\omega t} + \cdots\right]. \end{aligned}$$

The effect of non-linearity is to introduce harmonics $2f$, $3f$, ... nf as well as the fundamental frequency f itself. The amplitude distortion index is given by

$$\delta = \frac{\sqrt{(a_2^2 + a_3^2 + a_4^2 + \cdots + a_n^2)}}{a_1}. \tag{IV.24}$$

When the signal E consists of several sinusoidal functions of frequencies $f_1, f_2, \ldots, f_m$, the signal at the output becomes

$$\begin{aligned} S = {} & a_1 e^{j(\omega_1+\omega_2+\cdots\omega_m)t} \\ & + a_2 e^{j^2(\omega_1+\omega_2+\cdots\omega_m)t} \\ & + \cdots \\ & + a_n e^{j^n(\omega_1+\omega_2+\cdots\omega_m)t}, \end{aligned}$$

which can also be written in the form

$$S = a_1 \sum_i \sin \omega_i t + a_2 \sum_{i,j} (\sin \omega_i t \sin \omega_j t) + \cdots + a_n \sum_{i,j \ldots k} (\sin \omega_i t \sin \omega_j t \cdots \sin \omega_k t). \quad \text{(IV.25)}$$

When the input signal is of a composite nature, the amplitude non-linearities of the transfer network have the result of producing at the output not only harmonics of the fundamental frequencies, but linear combinations of the latter. This phenomenon will be studied in more detail at the beginning of the section on cross-modulation in multiplex systems.

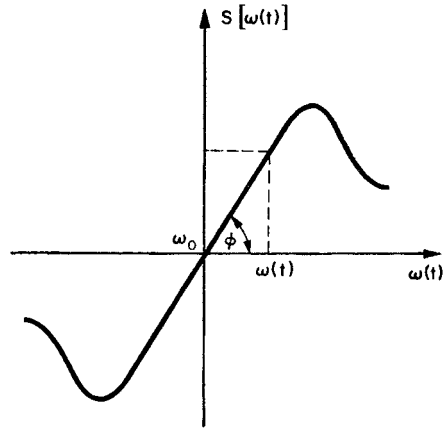


FIG. IV.19. Detection characteristic of an FM discriminator.

In a transmission channel, the modulation and demodulation devices are non-linear elements, and are the principal sources of amplitude distortion.

The distortion produced by an FM discriminator. Expanding the signal which energes from the discriminator, $S[\omega(t)]$, as a Taylor series, we obtain

$$S[\omega(t)] = S(\omega_0) + \frac{\omega - \omega_0}{1!} S'(\omega_0) + \frac{(\omega - \omega_0)^2}{2!} S''(\omega_0) + \frac{(\omega - \omega_0)^2}{3!} S'''(\omega_0) + \cdots.$$

Assuming the FM signal to be sinusoidal, the instantaneous frequency is of the form

$$\omega(t) = \omega_0 + \Delta\omega \sin \Omega t,$$

so that

$$S[\omega(t)] = S(\omega_0) + \Delta\omega \sin \Omega t\, S'(\omega_0) + \frac{\Delta\omega^2 \sin^2 \Omega t}{2} S''(\omega_0) + \frac{\Delta\omega^3 \sin^3 \Omega t}{6} S'''(\omega_0) + \cdots.$$

The discriminator characteristic is illustrated in Fig. IV.19.

If the discriminator is symmetrical, which is generally the case, we have

$$S(\omega_0) = O$$

$$S'(\omega_0) = p$$

$$S''(\omega_0) = A$$

$$S'''(\omega_0) = B$$

in which p is the slope at the origin of the discriminator ($p = \tan \varphi$).

The amplitude of the fundamental term is $S_o = p\,\Delta\omega$, and so

$$S[\omega(t)] = S_0 \sin \Omega t + \frac{A}{2} (\Delta\omega)^2 \frac{1 - \cos 2\Omega t}{2} + \frac{B}{6} (\Delta\omega)^3 \left[\frac{3}{2} \sin \Omega t - \frac{3}{4} \sin 2\Omega t - \frac{1}{4} \sin 3\Omega t\right]$$

$$= \left[S_0 + \frac{B}{4} (\Delta\omega)^3\right] \sin \Omega t - \frac{A}{4} (\Delta\omega)^2 \cos 2\Omega t - \frac{B}{4} (\Delta\omega)^2 \sin 2\Omega t - \frac{B}{24} (\Delta\omega)^3 \sin 3\Omega t + \frac{A}{4} (\Delta\omega)^2.$$

The term $A(\Delta\omega)^2/4$ is steady, and hence easy to detect. We write $\Delta S_c = A(\Delta\omega)^2/4$, which is also equal to the coefficient of the term in $\cos 2\Omega t$. We denote the weighting of the third harmonic by K_3:

$$K_3 = \frac{B}{24} (\Delta\omega)^3 \cdot \frac{1}{S_0}$$

and that of the second harmonic by K_2:

$$K_2 = \sqrt{\left[\left(\frac{\Delta S_c}{S_0}\right)^2 + \left(\frac{6K_3}{\Delta\omega}\right)^2\right]}.$$

In fact, the ratio $6K_3/\Delta\omega$ is far smaller than $\Delta S_c/S_0$ so that

$$K_2 \simeq \frac{\Delta S_c}{S_0},$$

and

$$K_3 = \frac{B(\Delta\omega)^3}{24} \cdot \frac{1}{S_0}.$$

IV.4.2. *Phase Distortion*

As already seen in equation (IV.22) the phase distortion is associated with the function $\psi(\omega)$.

If $e^{j\omega t}$ is the signal at the input of a network with a transfer function $A(\omega)\, e^{j\psi(\omega)}$, the signal at the output is of the form

$$A(\omega)\, e^{j[\psi(\omega)+\omega t]}.$$

The angular frequency of the output thus becomes

$$\omega_s = \omega + \frac{d\psi(\omega)}{dt}$$

or

$$\omega_s = \omega + \frac{d\psi(\omega)}{d\omega} \cdot \frac{d\omega}{dt}, \tag{IV.26}$$

thus $d\psi(\omega)/d\omega = \tau$ represents the group phase delay through the network.

The causes of distortion due to non-linearity of the phase characteristic lie essentially in the variation of the phase delay in amplifiers and filters, in reflections in transmission lines, and in propagation through several channels.

IV.4.2.1. *Variation of Phase Delay in Amplifiers and Filters*

In all RF or IF amplifiers, a resonant circuit forms the anode load of a valve or of the collector of a transistor.

For a simple tuned circuit, the potential across the terminals can be written in the form

$$\frac{V(\omega)}{V(\omega_0)} = \frac{1}{1 + jQ_0\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)}.$$

Using the above notation, the phase shift introduced by this circuit is

$$\psi(\omega) = -\tan^{-1}\left[Q_0\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right].$$

Differentiating this expression with respect to ω, gives the phase delay

$$\tau = \frac{d\psi(\omega)}{d\omega} = \frac{1}{1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2} \times Q_0\left(\frac{1}{\omega_0} + \frac{\omega_0}{\omega^2}\right).$$

In general, the frequencies encountered are near the central frequency ω_0, and so

$$\tau = \frac{2Q_0}{\omega_0\left[1 + Q_0^2\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)^2\right]} \simeq \frac{2Q_0}{\omega_0\left[1 + Q_0^2\left(\frac{\Delta f}{f_0}\right)^2\right]}$$

$$\Delta f/f_0 = 2(f - f_0)/f_0.$$

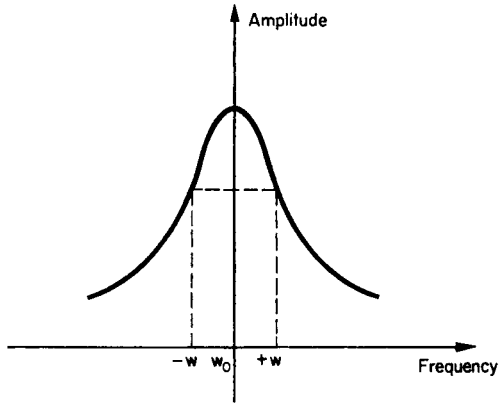


FIG. IV.20. Response characteristic of a band-pass filter.

The phase delay has its maximum value, τ_0, when $\omega = \omega_0$, that is at the peak of the response curve of the band-pass filter, as seen in Fig. IV.20, and

$$\tau_0 = \frac{2Q_0}{\omega_0}.$$

The 3 db bandwidth, $\mathscr{B}$, is such that

$$Q_0 \frac{\Delta f}{f_0} = 1 \quad \text{with} \quad \Delta f = \mathscr{B}.$$

The phase delay takes the value

$$\tau_{3db} = \tau_0/2,$$

and the variation of the phase delay within the band is then

$$\Delta\tau = \frac{\tau_0}{2} = \frac{Q_0}{\omega_0}.$$

In a wide band amplifier a succession of staggered circuits can be employed, as in a "triplet" unit. The variation of the phase delay in the pass-band then becomes

$$\Delta\tau = \frac{0{\cdot}22}{\mathscr{B}}$$

and for n triplets in cascade, we find

$$\Delta\tau = n\,\frac{0{\cdot}22}{\mathscr{B}}.$$

The influence of phase delay on multiplier stages. Another requirement is the choice of the circuits for the different multiplier stages. The distortions of order 3 are in fact directly proportional to the variation of phase delay in the band. For a circuit with tuned primary and secondary, the variation $\Delta\tau$ is given by

$$\Delta\tau = \frac{4\sqrt{2}}{\pi}\,\frac{\Delta f^2}{\mathscr{B}^3},$$

Δf is the frequency deviation in FM and $\mathscr{B}$ is the width of the pass-band at 3 db of the circuit in question.

In general, it is preferable that the distortions at the various stages should be the same; this is equivalent to requiring that the phase delay should be equal at each stage. From this we can deduce the requirement

$$\frac{\Delta f^2}{\mathscr{B}_1^3} = \frac{(k\Delta f)^2}{\mathscr{B}_k{}^3}$$

in which k is the multiplication constant, and $\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_k$ are the bandwidths associated with the successive multiplier circuits. A relation which links the bandwidths between successive stages can be obtained:

$$\mathscr{B}_k = k^{2/3}\mathscr{B}_1.$$

Phase distortion at the output of a frequency discriminator. The signal appearing at the output of the discriminator is

$$S = \frac{\mathrm{d}}{\mathrm{d}t}\left[\omega_0 t + \varphi + \psi\left(\omega_0 + \frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)\right]$$

$$= \omega_0 + \frac{\mathrm{d}\varphi}{\mathrm{d}t} + \frac{\mathrm{d}}{\mathrm{d}t}\left[\psi\left(\omega_0 + \frac{\mathrm{d}\varphi}{\mathrm{d}t}\right)\right].$$

On expanding the function as a Taylor series, and assuming the discriminator to be symmetrical, $\left[\omega_0 = 0 \text{ and } \left(\frac{d\psi}{d\omega}\right)_{\omega-\omega_0} = 0\right]$ we obtain

$$S = \frac{d\varphi}{dt} + \frac{1}{2!}\left(\frac{d^2\psi}{d^2\omega}\right)_{\omega=\omega_0} \cdot \left(\frac{d^2\varphi}{dt^2}\right)^2 + \frac{1}{3!}\left(\frac{d^3\psi}{d\omega^3}\right)_{\omega=\omega_0} \cdot \left(\frac{d^2\varphi}{dt^2}\right)^2 + \cdots.$$

Suppose that the modulation signal is sinusoidal, $s(t) = \sin \Omega t$, so that

$$\frac{d\varphi}{dt} = \Delta\omega \sin \Omega t;$$

the weightings associated with the 2nd and 3rd harmonics are then

$$K_2 = \frac{\Omega\Delta\omega}{2}\left(\frac{d\tau}{d\omega}\right)_{\omega=\omega_0}$$

and

$$K_3 = \frac{\Omega\Delta\omega^2}{4}\left(\frac{d^2\tau}{d\omega^2}\right)_{\omega=\omega_0}$$

respectively.

In practice, a discriminator has a point of inflexion at the origin (for $\omega = \omega_0$), and so

$$\frac{d^2\psi}{d\omega^2} = \frac{d\tau}{d\omega} = 0.$$

This means that in most cases, the weighting associated with the 3rd harmonic of the distortion will dominate.

IV.4.2.2. *Variation of Phase Delay in Transmission Lines*

In transmission lines, just as in aerial feeders, cavity filters and waveguides, mis-matching of the terminating impedances results in phase distortion. If V_e and V_s are the voltages at the input and output of a transmission line of length l, we have the following relationships:

$$V_e = V_s e^{j\varphi},$$

$$\varphi = \frac{\omega l}{v} = \frac{2\pi l}{\pi},$$

$$v = c/\sqrt{\varepsilon},$$

in which v is the propagation velocity in a dielectric with dielectric constant ε. Suppose a signal $e^{j\omega t}$ is applied to the input of the line whose reflection coefficients at the two ends, are ϱ_1 and ϱ_2. If τ is the delay for the whole line, the delay after reflection is therefore 2τ. (In this analysis subsequent echoes are neglected in comparison with the first.) The resultant signal at the output is then equal to

$$v_s = e^{j(\omega t+\varphi)} + \varrho_1\varrho_2\, e^{j[\omega(t-2\tau)+\varphi(t-2\tau)]},$$

in which

$$\tau = \left(\frac{d\varphi}{d\omega}\right)_{\omega=\omega_0}.$$

Further, neglecting the terms in τ^n $(n > 1)$:

$$\varphi(t - 2\tau) \simeq \varphi(t) - \frac{2\tau}{1!}\frac{d\varphi}{dt}$$

or

$$v_s = e^{j(\omega t+\varphi)} + \varrho_1\varrho_2\, e^{j[\omega t+\varphi(t)]}\, e^{j(2\omega\tau-2\tau d\varphi/dt)}$$

$$= e^{j(\omega t+\varphi)}[1 + \varrho_1\varrho_2\, e^{2j\tau(\omega-d\varphi/dt)}].$$

The distortion term, $D = \varrho_1\varrho_2\, e^{2j\tau(\omega-d\varphi/dt)}$, is proportional to the product $\varrho_1\varrho_2$ of the reflection coefficients. For a lossy line, $\varrho_1\varrho_2$ must be replaced by $\varrho_1\varrho_2\, e^{-2\alpha l}$.

The variation of the phase delay is

$$\Delta\tau = \frac{d\varphi}{d\omega} + \frac{d}{d\omega}(\tan^{-1}\theta)$$

with

$$\tan\theta = \frac{\varrho_1\varrho_2 \sin[2\tau(\omega - d\varphi/dt)]}{L + \varrho_1\varrho_2 \cos[2\tau(\omega - d\varphi/dt)]}.$$

Given that the phase delay varies between two extreme values, its value is given approximately by the following expression[32]

$$(\Delta\tau)_{max} = 4\varrho_1\varrho_2\frac{l}{v}.$$

IV.4.2.3. *Variation of the Phase Delay by Reflection*

Let d_1 and $d_2 = d_1' + d_2''$ represent the direct and indirect paths from A to B (see Fig. IV.21). The direct wave received at B is

$$V_1 = e^{j\varphi_1}$$

$$\varphi_1 = \frac{2\pi d_1}{\lambda} = \frac{\omega d_1}{c},$$

whereas the indirect wave is

$$V_2 = \varrho\, e^{-j\varphi_2}$$

$$\varphi_2 = \frac{2\pi d_2}{\lambda} + \pi = \frac{\omega d_2}{c} + \pi,$$

(the additional phase change of π is a consequence of the reflection). The reflection coefficient ϱ is clearly smaller than unity—it varies according to the nature of the terrain.

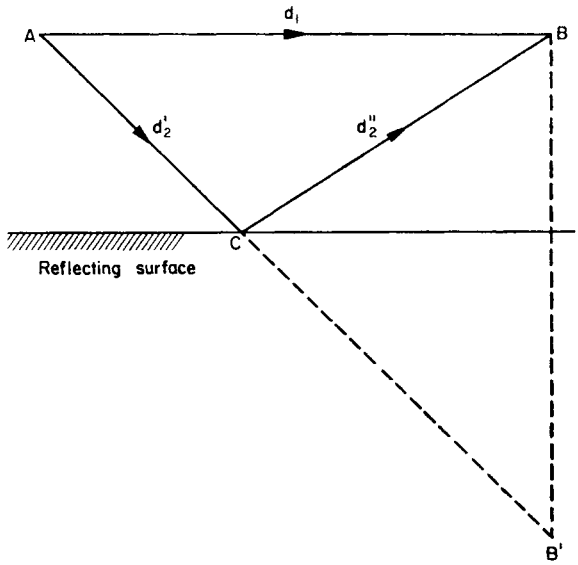


FIG. IV.21. Direct and reflected paths.

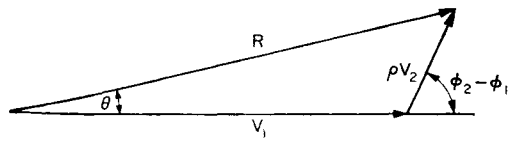


FIG. IV.22. Direct and reflected vectors.

The two vectors V_1 and V_2 combine (see Fig. IV.22) to give a resultant R. Applying the sine rule, we obtain

$$\frac{\sin\theta}{\varrho} = \frac{\sin(\varphi_1 - \varphi_2 - \theta)}{1}.$$

Whence

$$\tan \theta = \frac{\varrho \sin (\varphi_1 - \varphi_2)}{1 + \varrho \cos (\varphi_1 - \varphi_2)}.$$

The phase delay $\tau = \mathrm{d}\theta/\mathrm{d}t$ can easily be calculated as

$$\tau = \frac{\varrho[\varrho + \cos (\varphi_1 - \varphi_2)]}{1 + \varrho^2 \sin^2 (\varphi_1 - \varphi_2)} \cdot \frac{d_1 - d_2}{c}.$$

The delay τ varies between two extreme values corresponding to

$$\varphi_1 - \varphi_2 = +\pi \quad \text{and} \quad \varphi_1 - \varphi_2 = -\pi.$$

The maximum variation in the phase delay is therefore

$$(\Delta\tau)_{\max} \simeq \frac{|2\varrho(d_1 - d_2)|}{c}$$

where the reflection coefficient has been assumed to be small in comparison with unity.

CHAPTER V

FIELD PROPAGATION AND AERIALS ON MISSILES

A RADIO link consists of a transmitting aerial, a receiving aerial and the medium between them. Radio wave propagation is affected by various factors, and these are gathered together in this chapter. Missile technology involves rather special types of aerial, and it is with missiles in mind that we shall consider the problems associated with aerials.

V.1. GENERAL REMARKS

V.1.1. *The Gain, the Effective Area of an Aerial, and the Attenuation during Propagation*

Let $P(\theta, \varphi)$ be the power radiated in the direction θ, φ by an aerial, and let P_t be the total mean power radiated by the aerial; the gain is then defined by the ratio:

$$G(\theta, \varphi) = 4\pi \frac{P(\theta, \varphi)}{P_t}.$$

The "4" occurs because the two powers are compared per unit solid angle.

The gain, $G(\theta, \varphi)$ attains its maximum value for $\theta = \theta_0$ and $\varphi = \varphi_0$ $[G_{max} = G(\theta_0, \varphi_0)]$. An expression for the radiation diagram of the aerial is of the form:

$$\varrho = \frac{G(\theta, \varphi)}{G(\theta_0, \varphi_0)} = \frac{G(\theta, \varphi)}{G_{max}}.$$

The "effective area" of an aerial is defined as the ratio of the power received by the aerial to the power density per unit surface area in the region in which the aerial is placed. It can be shown that if S denotes the effective area, the gain, G, and S are related by

$$G = \frac{4\pi S}{\lambda^2}.$$

The attenuation of the transmission in free space at a distance d is of the form

$$A = \frac{P_E}{P_R} = \frac{(4\pi)^2 d^2}{\lambda^2 G_E G_R} = \frac{\lambda^2 d}{S_E S_R},$$

in which P_E and P_R are the powers at the transmitter and receiver respectively, and G_E and G_R, S_E and S_R are the gains and equivalent areas of the aerials at the transmitter and receiver respectively.

If Σ is the cross-sectional area of a paraboloid of diameter D, the effective area is given by $S = \alpha\Sigma(\alpha < 1)$ and the angular aperture by $(\Delta\theta)_{\max} = \frac{4}{\pi}\frac{\lambda}{D\sqrt{\alpha}}$. The efficiency of the aerial is in this case equal to 0·6.

The gain of a dipole, compared with that of an isotropic source, is $G = \frac{3}{2}$ (or 1·7 db) and the effective area is $S = 3\lambda^2/8\pi$.

The attenuation A which is quoted is valid for a given direction only, as the gains G_E and G_R are themselves functions of angles determined by the directional behaviour of the aerials. For omnidirectional aerials, we shall write $G_E = G_R = 1$.

When we are considering an ordinary line-of-sight link the attenuation can to a first approximation be regarded as that due to free space. Even under line-of-sight conditions, we have to consider the limitations imposed by the curvature of the earth.

If we assume that the refractive index of the earth's atmosphere varies linearly with altitude (the standard atmospheric approximation), it can be shown that the apparent radius of the earth is $\frac{4}{3}$ the real radius:

$$R_{\text{apparent}} = \tfrac{4}{3} R_{\text{real}}.$$

Three cases are possible, depending upon the type of missile employed.

(1) the missile may travel at a low altitude in the tropospheric zone
(2) the missile may travel at a great altitude, in the ionospheric zone
(3) the missile may cross all the atmospheric layers, and travel through the medium beyond the earth's atmosphere.

In addition the propagation is perturbed to a greater or lesser extent according to the frequency employed.

V.2. LOW ALTITUDE MISSILES

The missile receives two waves—one being a direct, line-of-sight wave, and the other being a wave which has been reflected from the ground, and which interferes with the first wave.

V.2.1. *Reflection From the Ground*

The ground from which the wave is reflected has a reflection coefficient, $\varrho = r\,e^{j\omega}$, which depends upon its complex dielectric constant. This dielectric constant varies widely with the surface which may be, for example, desert soil, a sheet of salt water or terrain covered with vegetation.

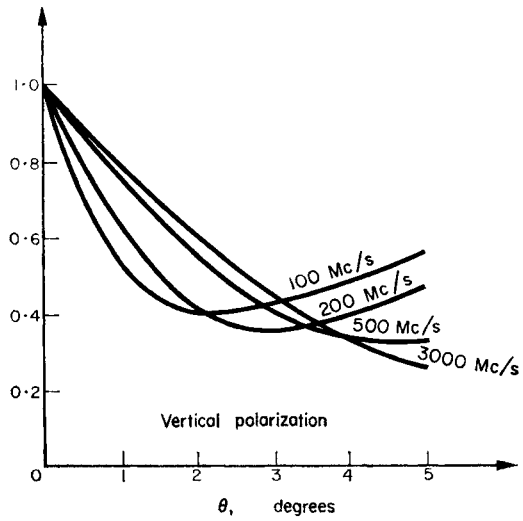


FIG. V.1. Variation of reflection coefficient with angle of incidence.[1]

In turn, the dielectric constant of the ground varies with the wavelength of the signal which is reflected from it.

Finally, the reflection coefficient is also a function of the angle of incidence, as can be seen from the curves plotted in Figs. V.1 and V.2.

For horizontally polarized waves, the coefficient r is equal to unity, and the phase angle is equal to the constant value π.

If V_0 is the amplitude of the direct wave, the amplitude of the wave reflected from the ground will be

$$V = V_0 \sqrt{(1 + r^2 + 2r \cos \varphi)}.$$

When the angle of incidence, φ, is small, we have

$$V = V_0 \frac{4\pi h_1 h_2}{\lambda d}$$

in which h_1 and h_2 are heights of the two aerials above ground level, and d is the distance between these two aerials.

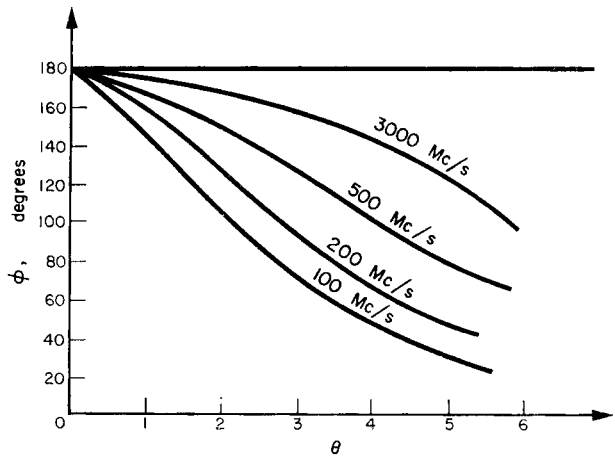


FIG. V.2. Variation of the argument of the reflection coefficient with angle of incidence.

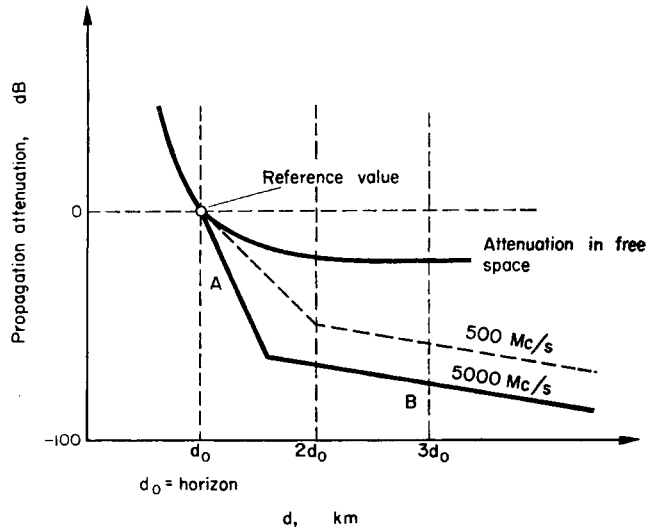


FIG. V.3. Variation of the propagation attenuation as a function of distance.[17]

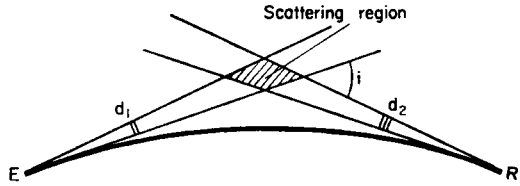


FIG. V.4. Localization of the scattering region.

V.2.2. *Tropospheric Propagation*

It is found that with powerful UHF transmitters, perfectly stable fields can be received beyond the horizon. Plotting attenuation as a function of distance (see Fig. V.3) the attenuation beyond the horizon is found to consist of two types in the regions (*A*) and (*B*). That in (*A*) is due to diffraction and in (*B*) to scattering.

The scattering phenomenon can be regarded as arising from dielectric discontinuities of the atmosphere. It occurs in the scattering region at the intersection of the beams d_1 and d_2 shown in Fig. V.4.

Another theory is based upon the idea that a series of refractions occurs when the beam crosses the troposphere, due to the lamination of the atmosphere. It is known that the radio energy is scattered in the direction of propagation by turbulence, and it is an experimental fact that this scattered energy decreases when the scattering angle and the angles of incidence, i, increase.

The supplementary attenuation of the transmission (which appears over and above the free space attenuation) varies slightly with frequency.

The slope of the characteristic in region (*A*) depends upon the frequency, while that in (*B*) is independent of it.

From Fig. V.3 it can be seen that the slope of the supplementary attenuation in the scattering zone, plotted as a function of distance, is very slight ($\sim$ 0·08 db/km). It should be noted that this field attenuation is a mean value; since the scattering has its origin in atmospheric turbulence, the instantaneous value of the attenuation varies continuously, daily and seasonally. The daily changes, which are of particular interest to us, are $\sim \pm 20$ db in magnitude. This relative fluctuation falls as the distance increases.

There is a relation between attenuation and the refractive index n which has an important influence on tropospheric propagation.

In Fig. V.5 are plotted the curves which express the normalized refractive index, $N = (n - 1) \times 10^6$, as a function of the height above ground level, h. Curve (1) represents the gradient of the refractive index in the case of total reflection (the wave falls at grazing incidence on the layer which has the refractive index in question). Curve (2) shows the normal gradient in the ordinary atmosphere, and curve (3) corresponds to an intermediate gradient.

The refractive index depends upon several parameters:

(1) the absolute temperature T

(2) the pressure p (measured in millibars)

(3) the saturated vapour pressure, e.

N has been found to satisfy the following empirical rule:

$$N = \frac{79}{T}\left[p - e + \frac{4800e}{T}\right].$$

The supplementary losses which have to be added to the attenuation of the transmission in free space are then those shown in Fig. V.6. The smaller the refractive index, the greater is the importance of these losses.

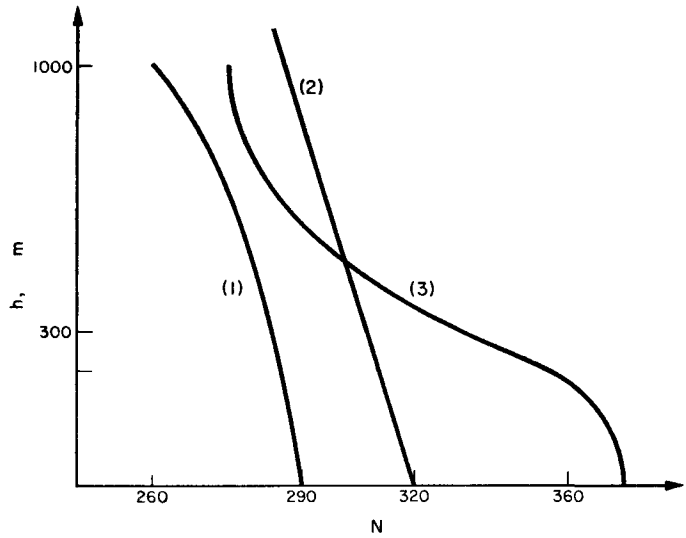


FIG. V.5. Variation of refractive index N as a function of height h above ground level.[17]

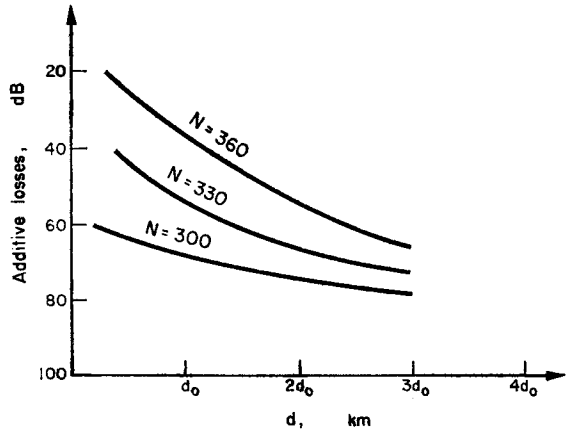


FIG. V.6. Supplementary losses for various values of refractive index.

V.3. MISSILES TRAVELLING AT GREAT ALTITUDES

The ionosphere consists of a region which has been ionized by the u.v. rays of the upper atmosphere; this region is divided into layers, the altitudes of which vary with time.

The absorption of radio waves in the ionosphere can be characterized by a viscous resistance, proportional to the velocity of the free electrons, v. To a first approximation, the ionosphere can be regarded as a stratified medium in which the electron concentration is proportional to the altitude. The equations which describe the wave propagation then become:

$$\text{curl}\,\mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \varrho\mathbf{v}$$

$$\text{curl}\,\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

in which $\mathbf{E}, \mathbf{H}, \mathbf{D}, \mathbf{B}$, and ϱ are the electric field intensity, the magnetic field intensity, the electric flux density, the magnetic flux density and the space charge density respectively; $\mathbf{v}$ is the velocity of the charges. If the medium is unpolarized,

$$\mathbf{D} = \varepsilon_0\mathbf{E}$$

$$\mathbf{B} = \mu_0\mathbf{H}$$

For sinusoidal waves, we have

$$\text{curl}\,\mathbf{H} = \mathrm{j}\omega\varepsilon_0\mathbf{E} + \varrho\mathbf{v}$$

$$\text{curl}\,\mathbf{E} = -j\omega\mu_0\mathbf{H}.$$

Consider an electron, mass m, moving with velocity $\mathbf{v}$, and subjected to a viscous force $\mathbf{f} = -r\mathbf{v}$ by the medium (r is the coefficient of viscosity).

We write

$$\beta = r/m$$

$$\varepsilon_0\mu_0 = 1/c^2$$

$$\omega/c = 2\pi/\lambda = k$$

$$\varrho = Ne$$

$$\omega_0^2 = \mu_0 c^2 Ne^2/m$$

$$\omega_0^2 = \alpha z \qquad \text{(the electron concentration is proportional to the altitude).}$$

The propagation equation becomes

$$\frac{\mathrm{d}^2E}{\mathrm{d}z^2} - \frac{\omega^2}{c^2}\left[\frac{\alpha z}{\omega^2\left(1 - \dfrac{j\beta}{\omega}\right)} - 1\right]E = 0.$$

It can be shown[69] that if the gradient of the electron concentration remains constant, the reflection coefficient takes the form

$$|\varrho| = \exp\left[-4\beta\omega^2/(3\alpha c)\right]$$

which results in an attenuation of the reflected wave of

$$\frac{4\beta\omega^2 \times 8{\cdot}68}{3\alpha c} \text{ db.}$$

The coefficient α is related to the refractive index n by

$$n = \sqrt{(1 - \alpha z/\omega^2)}.$$

The altitude z is measured from the base of the ionospheric layer. It can be seen that the refractive index is a function not only of altitude but also of the frequency of the incident wave. Further, for a given altitude and refractive index, reflection of the incident wave plays a more important role when the frequency is low. A low frequency wave finds it more difficult to penetrate the ionospheric layers than does a wave with a higher frequency. The Russian sputniks transmit at about 20 Mc/s, and are as a result harder to follow by radio than the American missiles which use a higher frequency.

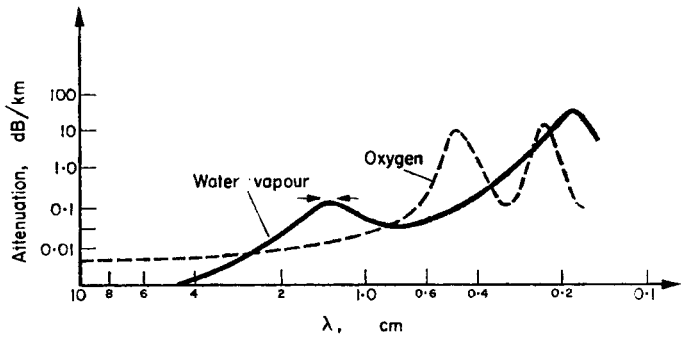


FIG. V.7. Molecular attenuation as a function of wavelength.

In the preceding calculations, it has been assumed that the ionospheric medium possesses a certain degree of homogeneity; in fact, this is not so except within laminar zones or superimposed layers. This laminated structure of the atmosphere results in a certain curvature of the propagation, again due to variation of the refractive index; the curvature is defined by

$$\frac{1}{R} = \frac{1}{n}\frac{\mathrm{d}n}{\mathrm{d}s}$$

in which R is the radius of curvature.

It should also be noted that the radio waves are absorbed by the ionospheric medium, principally as a result of the presence of oxygen and water vapour. This phenomenon is independent of the polarization of the wave, but depends upon frequency: it is negligible for frequencies below 5000 Mc/s.

The process of absorption by oxygen is based upon the magnetic interaction between the radio wave and the paramagnetism of the oxygen molecule. Absorption by water vapour, on the other hand, is related rather to the electric polarization of the water molecule. For certain frequencies, the absorption is particularly abnormal, and this corresponds to molecular resonance of the medium at the frequency in question.

In Fig. V.7, the attenuation curve is plotted in db/km as a function of the wavelength.

At the wavelengths $\lambda = 0{\cdot}5$ and $0{\cdot}25$ cm, attenuation maxima are to be seen for oxygen, and when $\lambda = 1{\cdot}35$ cm, the attenuation is a maximum for water vapour.

V.4. VARIATIONS IN THE GAIN OF AERIALS

It has already been mentioned that it is less easy to obtain the effective values of gains of aerials carried on missiles compared with those measured in free space. The apparent gains of the aerials at the transmitter and the receiver depend not only upon one another but also upon the trajectory which the propagation has followed.

The fall in the gain of an aerial is due to atmospheric turbulence. The surface normal to the direction of the incident wave varies with the curvature of lines of equal field-strength which results from turbulence. This deformation of the wave surface is equivalent to a reduction of the gain of the aerial; and the smaller the equivalent surface area of the aerial, the smaller is the effective incident wave surface.

The turbulent phenomena are consequences of variations in the meteorological conditions; they depend upon the changes in the nature of the terrain between the transmitter and the receiver. Tropospheric links are particularly susceptible to this. For example, measurements on a 100 Mc/s link over a distance of 500 km, showed that for an aerial whose gain had been measured to be 30 db in free space, the loss was 2 db.

V.5. THE PROBLEM OF ATTACHING AERIALS TO A MISSILE

This aerial problem has to be considered in conjunction with the aerodynamic considerations, the use to which the missile is to be put, and as a result, with the field pattern.

When the missile has to be able to travel in any direction whatsoever, an omnidirectional aerial must be used. If, on the contrary, the missile in question is to move in a particular direction, and to remain within a cone, the apex of which is situated at the transmitter or receiver aerial on the ground, it is profitable to employ a certain degree of directivity. The shape of the pattern is then set by the departures which the missile is likely to make from the intended trajectory.

Consider the case of a tactical missile which is fired from a transmitting station E towards its target B. The aerial on the missile will have to radiate backwards, and emit as little radiation as possible in the direction of motion of the missile, in order to prevent detection by the enemy (see Fig. V.8).

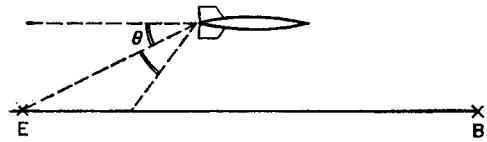


FIG. V.8. An aerial on a missile, radiating backwards.

If the angular aperture, θ, of the field pattern is too narrow, there is a risk of losing the missile should it depart suddenly from its forecast trajectory, as a result, for example, of a gust of wind across its path.

The choice of polarization of the wave is equally important. In turning, a missile may well bank at more than 60°; the effective height of the aerial or the resolved polarization is then multiplied by a factor smaller than unity (approximately equal to cos 60°). An appreciable loss in the gain ensues; if the missile should be far enough away, the radio link (for the remote steering of a rocket, for example) may well be interrupted and this could result in serious disturbance to the flight behaviour of the missile. The choice of polarization (horizontal, vertical or circular) depends upon the use and the trajectories foreseen for the missile.

V.5.1. *Classification of Aerials and their Electrical Parameters*

A standard way of classifying aerials is in terms of the frequencies which they radiate; this is shown in the following list:

LF: $f < 2$ Mc/s

HF: $2 < f < 20$ Mc/s

VHF: $20 < f < 250$ Mc/s

UHF: $250 < f < 1250$ Mc/s

SHF: $f > 1250$ Mc/s.

Although this classification is perhaps valid in a general way, it is very obvious that when the aerial is to be mounted on an aeroplane or a missile, the lateral dimensions of the latter cannot be ignored. The shorter is the wavelength employed in comparison with the dimensions of the aircraft, the closer does the real diagram resemble the theoretical diagram which would be obtained over an earthed plane of infinite extent. Taking these considerations into account, aerials can be separated into four categories, according to the wing-span "E" of the missile:

$$\lambda > 6E \tag{1}$$

$$\lambda/6 < E < 3\lambda \tag{2}$$

$$3\lambda < E < 20\lambda \tag{3}$$

$$E > 20\lambda. \tag{4}$$

The electrical parameters which define an aerial are

(i) the radiation diagram
(ii) the impedance
(iii) the maximum gain.

The relative importance of these various factors depends not only upon the frequency at which the aerial is to work, but also upon its application.

If the aerial is imperfectly matched to its feeder, a standing wave ratio different from unity results; the reflection coefficient $r = (s - 1)/(s + 1)$ is not zero. This is important since, as has been seen, the phase distortion in the feeder is proportional to the reflection coefficient and in addition there is a loss of power.

At the beginning of this chapter, the radiation diagram was defined by

$$\varrho = \frac{G(\theta, \varphi)}{G_{\max}}.$$

It must be ensured that, during the missile flight the radiation diagrams of the missile and the terrestrial station aerials are such that the electric field is never interrupted. Consider this further in terms of the example shown in Fig. V.9. Suppose that Ox, Oy and Oz are the coordinate axes of a fixed reference system. The terrestrial aerial lies along Oz. The axes associated with the aircraft or missile are GX, GY, GZ (these are the axes of rolling, pitching and yawing). The radiation diagrams of the terrestrial aerial and the missile aerial are well-defined surfaces, and depend upon the characteristics selected according to the function of the missile. If the direction OG should correspond to a gap, either in the terrestrial diagram or in the one in flight, the reception of the RF wave will be interrupted. This is equivalent to saying that the directions of the vector OG which have to be excluded are those for which the product $\varrho_1(\theta, \varphi)$, $\varrho_2(\theta, \varphi)$ is zero.

The problem can be simplified still further by regarding the two aerials—along Oz and OZ,—as dipoles; the diagrams are spherical (see Fig. V.10). For clarity's sake, a plane is considered (which is justified by the symmetry of the diagram). The form of the radiation diagram in the plane yOz is described by

$$\varrho_1(\theta, \varphi) = 30\pi I_0^2 \sin^2 \theta \frac{l^2}{\lambda^2}.$$

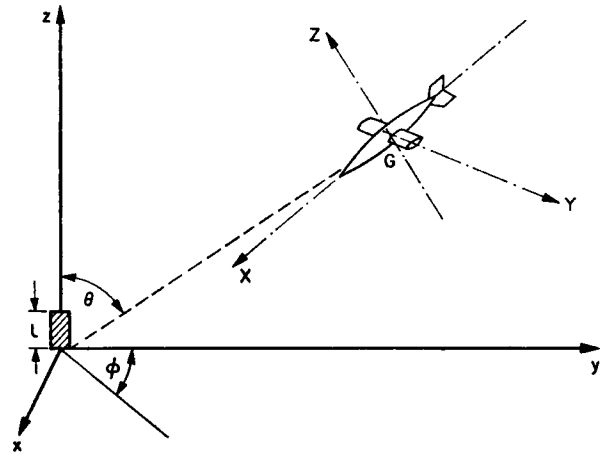


FIG. V.9. Fixed coordinate axes, and those defined by the missile.

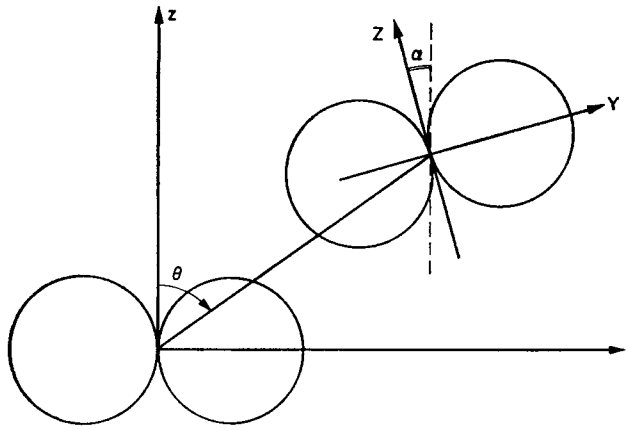


FIG. V.10. Relative positions of dipole polar- diagrams on the missile and on the ground.

The diagram relative to the missile can be deduced by replacing θ by $\theta + \alpha$ in this expression which gives

$$\varrho_2(\theta, \varphi) = 30\pi I_0^2 \sin^2(\theta + \alpha) \frac{l^2}{\lambda^2}.$$

The product of these two functions for the diagrams is thus of the form:

$$\varrho_1\varrho_2 = k[\sin\theta \sin(\theta + \alpha)]^2,$$

and there are two directions for which the field will be cut off:

$$\theta = 0$$

and

$$\theta = -\alpha.$$

It should be noticed that we have made no distinction between the two aerials, and the reasoning is thus equally valid irrespective of whether the aerial on the missile be the transmitter or receiver (telemetry or remote-control, respectively).

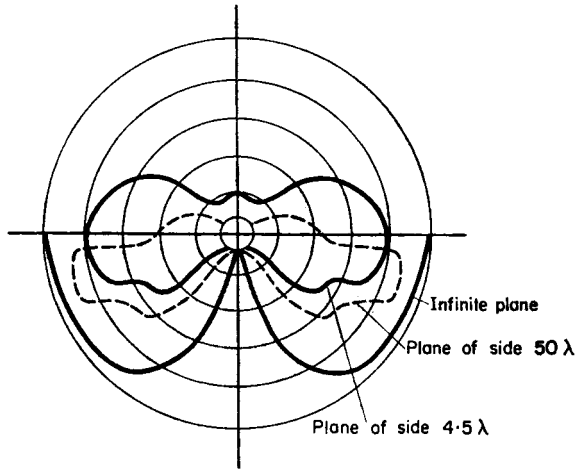


FIG. V.11. Deformation of the polar diagram of a dipole, over an earthed plane of various dimensions.

V.5.2. *The Influence of the Supporting Structure on Aerial Performance*

The radiation diagram of a dipole, placed over an earthed plane of infinite dimensions is a circle and if the length of the sides of the earthed plane is progressively reduced the diagram is deformed in the way shown in Fig. V.11.

When the dimensions of the sheet become smaller than 5λ, parasitic lobes appear which can be regarded as secondary radiation. Further, when the aerial is placed a certain height above the missile the resulting diagram consists of the principal diagram together with the parasitic lobes produced by reflection and diffraction of the wave at the various parts of the missile.

It is easy to obtain the diagram theoretically by considering the principal aerial and its electrical images with respect to the various different parts of

the structure. The whole diagram is composed of the combined effect of these individual dipoles (the initial dipole together with its images). In particular the missile fuselage can be treated as a cylindrical conducting surface. The image of the aerial is then the aerial inverted with respect to the cylinder (a point P on the aerial, which lies at a radial distance OP from the axis of the cylinder, has its image inside the cylinder at the inverse point Q on OP such that $OP \cdot OQ = OR^2$, where OR is the radius of the cylinder; the complete image is built up by inverting all points on the aerial).

The parasitic diagram produced by diffraction results from the slender parts of the fuselage. The disturbance is particularly marked when the aerial is parallel to a ridge on the structure. Care should therefore be taken to place the aerials in such a way that they are perpendicular to any ridges.

In the case of propeller-driven missiles and aircraft, the rotation of the propeller modulates the HF wave; this parasitic modulation is particularly strong when the dimensions of the propeller are of the same order as the wavelength.

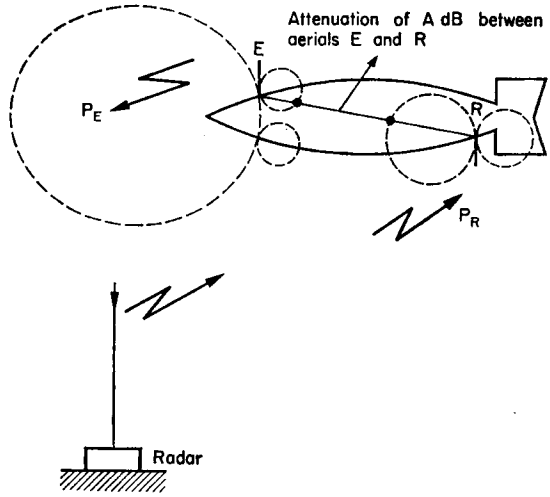


FIG. V.12. Principle of an echo reinforcer.

A target-missile (a remotely piloted aeroplane) carries a large number of radio instruments, each with its own well-defined function (telemetry, remote-control, measurement of altitude, measurement of the miss-distance, position-finding and echo-reinforcement for example). Each of these functions requires its own frequency, and in consequence, a separate aerial. When the frequencies are all very different from one another, and the aerials are far enough apart, the problem is easily solved. When this is not so the problem becomes extremely difficult. It is necessary to ensure beforehand that the coupling between the various aerials is sufficiently slight that the transmitters on the

missile do not jam the receivers. The echo-reinforcer is a particularly instructive case in point. This is an instrument which, on receiving a wave of frequency f, is designed to send out an echo at a frequency which is equal to, or very slightly different from f. Its purpose is to reinforce the power of the reflected echo; in its absence, the echo would be too weak for a tracking radar system to be able to detect the missile (the dimensions of a missile are usually smaller than those of an aircraft). This is equivalent to increasing the equivalent surface of the missile (see Fig. V.12).

Denoting by P_R and P_E the powers of the signals received and emitted by the echo reinforcer, which has a gain $G(G = P_E/P_R)$, the attenuation A of the wave emitted at the point of reception must be such that $P_R > AP_E$, which means that

$$|A \text{ db}| < |G \text{ db}|.$$

Consider a practical example: the equivalent surface of a bomber is about 60 m^2, whereas that of a target missile is 1 m^2. The gain of the echo-reinforcer will thus be 60, or about 20 db (= 10 log 60). The decoupling of the aerial will therefore have to be at least 30 db.

V.5.3. *Description of Some Types of Missile Aerials*

To avoid going outside the framework of this book the description of aerials is restricted to a few special types which are employed in connection with telemetry and remote-control. Standard types such as dipoles, truncated cones, Yagi aerials, helical aerials and disc-cones, which are already very well known, are not discussed.

V.5.3.1. *The Spiral Aerial*

As is well known the dimensions of an ordinary aerial impose a relatively narrow bandwidth on it. If an aerial could be defined in terms of its angular parameters alone, the performance of such an aerial would be independent of frequency. It is for this reason that a spiral aerial, defined solely in terms of angles, seems to be the ideal aerial, with a very wide bandwidth. In fact, there is a physical limitation in producing such an aerial, since the curvilinear abscissa must be kept within reasonable bounds. The bandwidth of such an aerial will, further be bounded by certain limits; values of the ratio f_{max}/f_{min} greater than ten can however be achieved. These aerials are employed for frequencies between 100 Mc/s and 10,000 Mc/s. Printed circuit technology is particularly well adapted for use in conjunction with these aerials.

Consider the case of a logarithmic spiral, of the form $\varrho = k\, e^{a\varphi}$. The radiation produced is a maximum in a direction normal to the plane of the spiral, and the radiation diagram is conical with a diameter of $\lambda/5$; the

angular aperture of the diagram is from 70° to 110°. Since the aerial is bi-directional, a matching cavity is required behind the spiral although its depth is not critical.

In the frequency range $f_{max} - f_{min}$, the impedance of the aerial varies from 40 to 70 Ω, and thus for a feeder of 50 Ω, a match is obtained with an SWR of, at worst, 1·4.

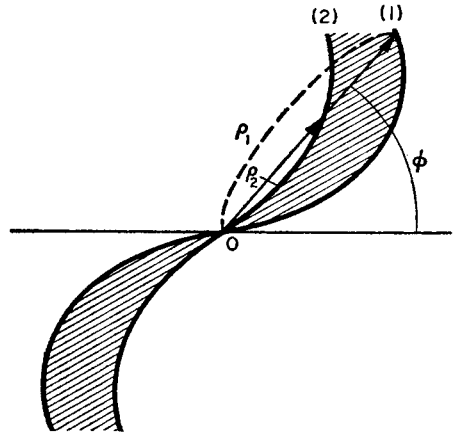


FIG. V.13. Spiral aerial.

The aerial is constructed in the manner illustrated in Fig. V.13; it consists of two arms, defined by

$$\varrho_1 = k\, e^{a\varphi}$$

$$\varrho_2 = k\, e^{a(\varphi-\alpha)}$$

and

$$\varrho_1 = k\, e^{a(\varphi-\pi)}$$

$$\varrho_2 = k\, e^{a(\varphi-\pi-\alpha)}.$$

Spiral (*2*) is derived from spiral *1* by a rotation of $\alpha°$ about the origin *0*. The second arm, which consists of spirals (*3*) and (*4*), is symmetrically placed about the origin with respect to the first arm.

If the mean length of the spiral is greater than the longest of the working wavelengths, the parameters are relatively independent of frequency. Good diagrams are obtained with spirals about one and a half turns long.

It is noteworthy that the polarization is rectilinear when the length of an arm is small in comparison with the wavelength. When the frequency is increased, it becomes elliptical, and finally circular. Further, polarization which is perfectly circular along the axis of maximum radiation becomes elliptical on either side of this direction.

V.5.3.2. *The Slot Aerial*

According to Babinet's principle, a dipole and a slot are inverse and complementary structures. The diagram of a rectangular window is therefore similar to that of a dipole (see Fig. V.14).

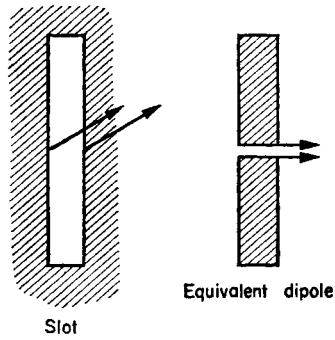


FIG. V.14. Analogy between a slot and a dipole.

The polarization of the waves is reversed; thus if the dipole provides vertical polarization, the complementary aperture will give horizontal polarization. The impedance of the slot is inversely proportional to that of a dipole,

$$Z_{\text{dipole}} \times Z_{\text{slot}} = (120\pi)^2.$$

The polarization which is provided in this way is therefore linear. The fading of the signal which results from the mismatching of the polarization during the missile's trajectory is a serious problem. To overcome this difficulty,

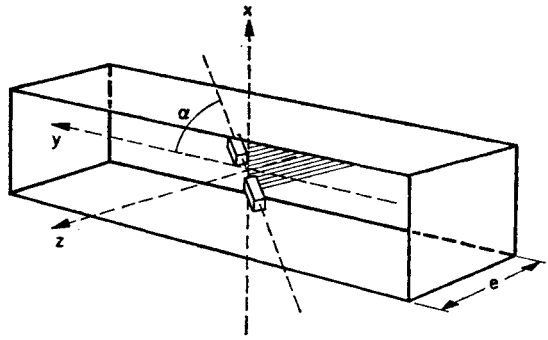


FIG. V.15. Excitation of slot aerial.

circular polarization must be employed either at the ground or on the projectile. If the polarization is made circular at the ground, there will be a loss of 3 db with respect to the maximum gain. It is, on the contrary, possible to obtain circular polarization on the missile by exciting the slot by means of a dipole, suitably inclined in the plane of the slot (see Fig. V.15).

When the dipole is inclined at an angle α, a transverse component appears which excites the slot; the phase of this component differs from that of the dipole current by about $\pi/2$. There are two fields, therefore, combining in quadrature; that of the dipole and that of the slot. If the centre of the dipole coincides with that of the aperture, a circularly polarized wave will be obtained.

V.5.3.3. *A Special Aerial Arrangement for Use with Spinning Missiles*

Consider the missile shown in Fig. V.16. At the tips of the two fins, 1 and 2, two folded dipoles are incorporated; these dipoles are denoted by D_1 and D_2.

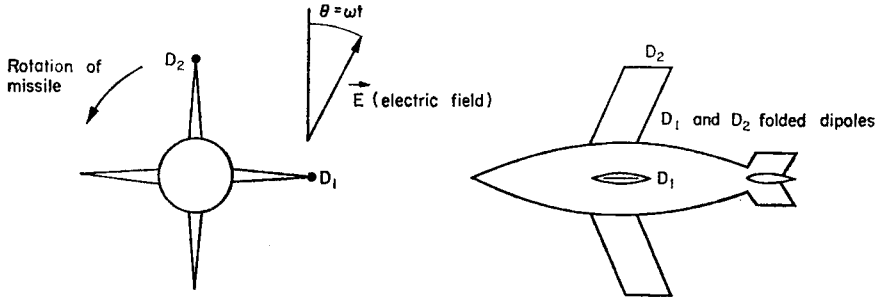


FIG. V.16. Aerials on a spinning missile.

The incident wave is of the form $Ee^{j\omega t}$. It falls simultaneously on the two aerials, and the amplitudes at D_1 and D_2 resepctively are

$$V_{D_1} = E \sin \theta = E \sin \omega t$$

$$V_{D_2} = E \cos \theta = E \cos \omega t.$$

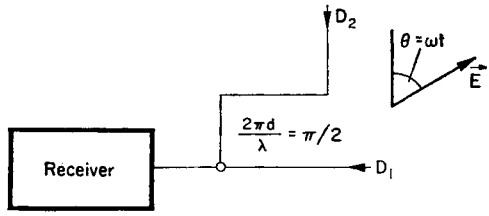


FIG. V.17. The phasing circuit.

If a phase-change of $\pi/2$ is introduced into one of the channels (see Fig. V.17), the resultant field is $2E \sin \omega t$.

The diagrams in the plane of the missile and in the plane normal to it are shown in Fig. V.18.

The degree of ellipticity is a function of the phase change $2\pi d/\lambda$ which is introduced.

An almost perfect match can be obtained with a standing wave ratio of 1·1 or 1·2 for the working frequency.

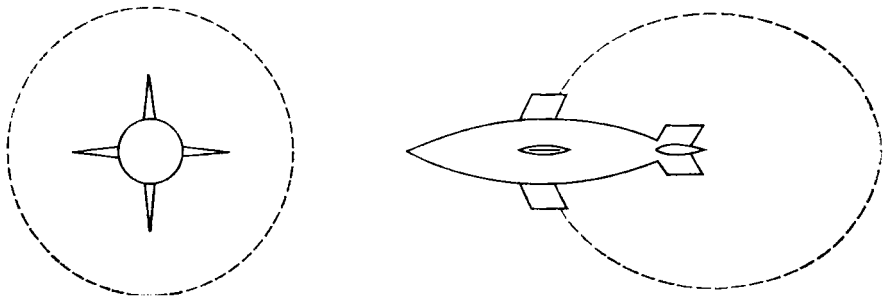


FIG. V.18. Radiation diagram in the plane of missile and plane normal to it.

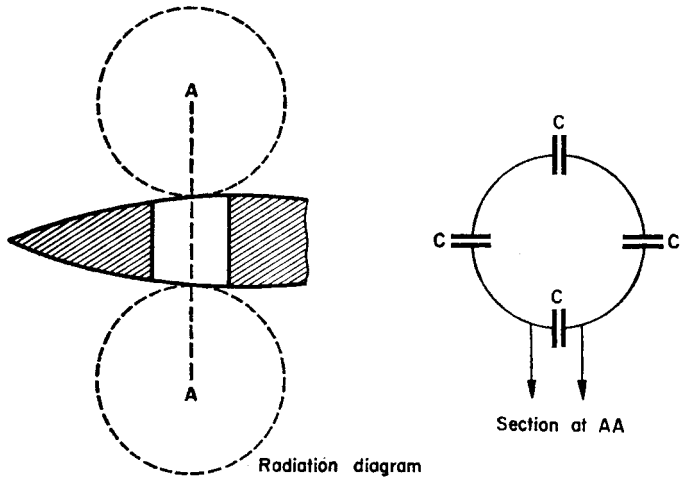


FIG. V.19. Layout of missile proximity-fuse aerial.

The diagrams of Fig. V.18 show that it is easy for a combat aircraft to follow a rotating missile.

It may be necessary to incorporate into this missile, which we suppose to be tactical, a device known as a proximity fuse; a device which causes the rocket to blow up as soon as it reaches a distance less than or equal to some predetermined critical value. We assume, in fact, that the missile never exactly reaches the goal. The aerial diagram should have symmetry of revolution about the axis of the missile, and the direction of maximum gain will be perpendicular to the missile itself. To achieve this, a Kandoian loop can be

employed, contained within an insulated section of the fuselage, in the way shown in Fig. V.19.

The radiation diagram consists of a torus, the inner part of the surface of which is tangent to the missile. The impedance characteristic of such an aerial is shown in Fig. V.20.

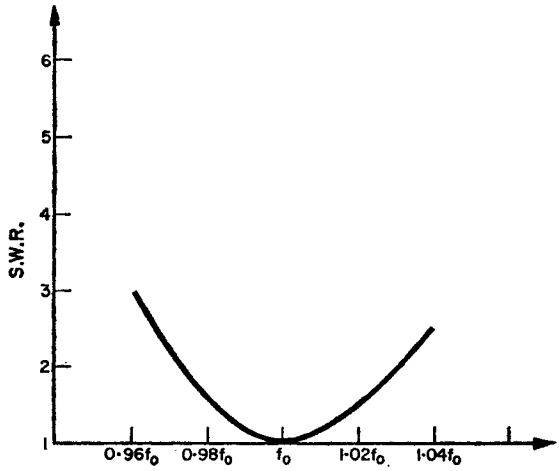


FIG. V.20. Variation of SWR as a function of frequency.

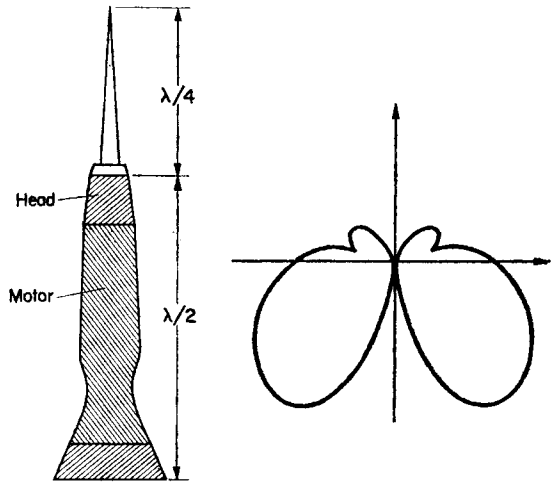


FIG. V.21. Description of the rocket-aerial.

This section concludes with a description of a telemetry aerial, designed to transmit back to earth information from a laboratory missile which might, for example, be used to study the influence of cosmic rays at high altitudes.

This rocket travels from low to high altitudes, therefore, and if the information is all to be received unimpaired at the ground, the aerial must radiate downwards, and with symmetry of revolution about the axis of the missile.

The aerial consists of the body of the missile itself (see Fig. V.21). The resonant rod ($\lambda/4$) placed on the nose of the missile excites the latter, which may be regarded as a dipole. If the diameter of the propulsion system is equal to one-fifth of the total length, the current distribution will be more uniform in the lower part of the dipole, and this results in a radiation diagram which is more intense in the downward direction. By varying the length of the rod, therefore, a situation can be reached in which the missile is perfectly excited.

V.6. PROCEDURES WHICH ARE EMPLOYED TO INCREASE THE EQUIVALENT AREA OF MISSILES STATICALLY

When a radar unit is being used to track a missile, it is usually found that the equivalent area of the missile is too small for the reflected echo to be received in the normal way by radar. This area is, therefore, artificially increased; some details are now given of how this may be done.

The "equivalent surface for a radar echo" can be defined as the surface-area of a section of a sphere which would reflect an echo of the same intensity as the system in question, in the direction of the incident radiation.

V.6.1. *Three Mutually Perpendicular Surfaces*

In optics, a light ray incident upon a trihedral reflector returns along a path parallel to the direction of incidence. This property remains valid at high frequencies (see Fig. V.22). A trihedral element of this kind has an angular aperture of $\pi/2$ steradians, so that for all directions in space to be covered, four such units would have to be used, each turned through $\pi/2$ radians with respect to its predecessor. In fact, only directions below a horizontal plane are of interest, and hence at most two of these trirectangular trihedral elements will be required.

The three perpendicular planes can be triangular, square, or circular (see Fig. V.23).

When the angle at the apex is $\pi/2$, the reflector is said to be "monostatic", that is, that the reflected ray is parallel to the incident ray.

Alternatively, trirectangular trihedra which have the property of reflecting the incident ray with a certain change in direction can be used. These reflectors are called "bistatic", and the apex angle of such trihedra varies between 87 and 93°.

The equivalent surface presented to a radar echo by a reflector is given by

$$\Gamma = A \times G$$

in which A and G are the equivalent surface and the gain respectively when the reflector is regarded as a radiating aerial. The relation between A and G is thus

$$G = \frac{4\pi A}{\lambda^2}.$$

For paraboloids, as we have already seen, the surface A is about 0·75, in which S is the actual surface area which the reflector presents.

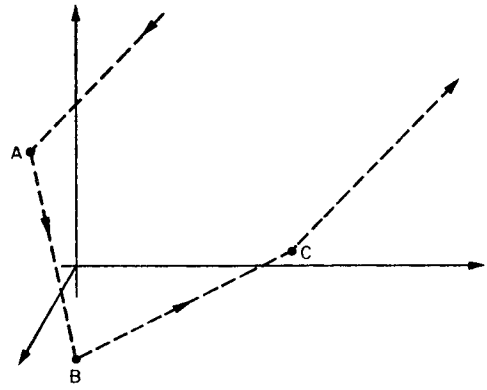


FIG. V.22. A trirectangular trihedron returns a reflected ray in the same direction as the incident ray.

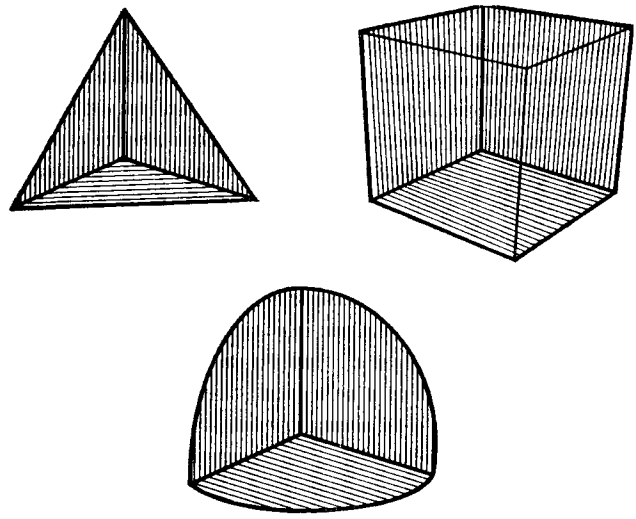


FIG. V.23. Common types of trirectangular trihedron.

To a first approximation, the calculations can be simplified by supposing that the relation

$$A = S$$

is true.

For a triangular, trirectangular trihedron, we have

$$S = \tfrac{3}{2}R^2 \cos\varphi$$

in which φ is the angle which the diagonal of the cube makes with an edge ($\varphi = 54{\cdot}74°$). Thus

$$S = 0{\cdot}866R^2.$$

In the case of a circular trirectangular trihedron, the surface area is

$$S = 1{\cdot}36R^2.$$

Further, the gain, G_{max}, is related to the angular aperture, α, of the radiation diagram by the formula

$$G_{max} = \frac{k}{\alpha^2}.$$

If α expressed is degrees, $k = 10{,}000$.

The gain of the monostatic reflector is greatest in the direction of the bisector of the apex angle of the corner. The gain falls by a half, with respect to its value along this direction, for directions which make an angle of 35° with the latter.

If the radius of the circle, r, in a circular trirectangular reflector is equal to at least λ (the wavelength), the equivalent surface is given by

$$\Gamma = \frac{16\pi}{3}\frac{r^4}{\lambda^2}.$$

The gain of such a reflector can be expressed in terms of a Gaussian curve; each point on the reflector is regarded as a source of radiation, the phases associated with the various points being assumed to be random.[40]

As a result

$$G = G_{max}\, e^{-\mu\left(\frac{\alpha}{\alpha_0}\right)^2},$$

with

$$G_{max} = \frac{k}{\alpha_0^2},$$

in which α_0 is the angle which corresponds to the gain at half the maximum power; for $\alpha = \alpha_0$, therefore,

$$G = \frac{1}{2}G_{max} = \frac{k}{\alpha_0^2}e^{-2\mu}$$

and hence

$$\mu = \tfrac{1}{2} \log_e 2 \simeq 0{\cdot}35.$$

The ratio of the equivalent surface for a radar echo, Γ, to the real surface, S, is of the form

$$\frac{\Gamma}{S} = \frac{k}{\alpha_0^2} e^{-2\mu\left(\frac{\alpha}{\alpha_0}\right)^2}.$$

The maximum value of α_0 for which the ratio Γ/S also takes its greatest value, with α fixed, can now be calculated:

$$\frac{d}{d\alpha_0}\left(\frac{\Gamma}{S}\right) = \frac{2}{\alpha_0^3} e^{-2\mu\left(\frac{\alpha}{\alpha_0}\right)^2}\left[\frac{2\mu\alpha^2}{\alpha_0^2} - 1\right].$$

Thus for $\alpha_0 = \alpha\sqrt{(2\mu)}$, Γ/S has a maximum:

$$\frac{\Gamma}{S} = 0{\cdot}5\frac{k}{\alpha^2} \quad \text{for} \quad \left(\frac{\alpha}{\alpha_0}\right) = 1{\cdot}201.$$

in which $k = 10{,}000$ if α is expressed in degrees.

V.6.2. *The Luneberg Sphere*

The Luneberg sphere is a dielectric lens. It consists of a family of concentric layers, the refractive index, n, of which varies according to the law

$$n = \sqrt{[2 - (\varrho/R)^2]}$$

in which ϱ is the distance between the point considered and the centre of the lens, and R is the maximum radius of the sphere.

The light rays follow the trajectories shown as dotted lines in Fig. V.24. The focus of such a lens lies on its own surface.

If a reflecting wall is placed on this surface it is found that an incident ray will, after reflection at the mirror, adopt the path which is symmetrically placed about the axis Ox with respect to the incident trajectory.

The function of echo reflection is thus efficiently performed for every direction which lies within a cone of angle θ. The maximum value of the angle θ which the lens will accept without suffering any loss of apparent surface, is equal to $\pi/2$; to be able to reflect in any direction, therefore, would require four lenses. For the region beneath a horizontal plane, two lenses suffice. However, a certain loss of equivalent surface can be tolerated, and in this case the maximum angle θ is 140°; which allows the use of a single Luneberg sphere.

The equivalent surface of a Luneberg lens is given by the expression

$$S = \frac{4\pi^3 R^4}{\lambda^2}$$

in which R and λ are the radius of the sphere and the wavelength employed, respectively.

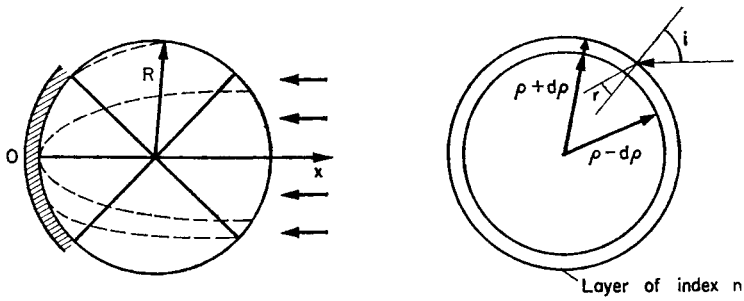


FIG. V.24. Luneberg sphere.

The equivalent surface is therefore a rapidly increasing function of frequency; the same sphere can, however, be used for frequencies over the range 3000 Mc/s to 10,000 Mc/s.

It should be mentioned that measurements of equivalent surfaces give results sensibly equal to half their theoretical values.

V.7. LINKS WITH DIVERSITY

The RF signals which reach a receiver fluctuate both in amplitude and in polarisation; the problem is to transmit the information along several independent transmission channels in parallel, in an effort to reduce the undesirable influence of noise or fading.

We shall distinguish between spatial diversity, polarization diversity, frequency diversity and phase diversity.

With spatial diversity, the aerials at the receiver are separated from one another by a certain distance in such a way that the aerials receive the same wave at different phases. If at a time t the amplitude of the resultant wave passes through a minimum at aerial number 1, it will have a higher value at aerial number 2.

For example, the phase difference between two aerials is given by

$$\Delta\varphi = \frac{4\pi h_E \cdot \Delta h}{\lambda d}$$

in which h_E is the height of the transmitting aerial
Δh is the distance between the two receiver aerials
d is the transmitter–receiver distance
λ is the wavelength.

Thus for

$$\Delta\varphi = \pi$$

we find

$$\Delta h = \frac{\lambda d}{4h_E}.$$

In fact, when the link is with a missile, d and h_E vary, and it is difficult to choose Δh in such a way as to arrive at $\Delta\varphi = \pi$ exactly, whatever the position of the missile.

With polarization diversity, the resultant signal comes from a combination of several signals which have been transmitted by carrier waves emitted with various different polarizations – horizontal, vertical, clockwise circular, and anti-clockwise circular; or again, these carrier waves may have elliptical polarizations, and be distinguishable only their different degrees of ellipticity.

When there is frequency diversity, the information in the signal is emitted simultaneously on carrier waves with different RF frequencies. The difference, Δf, between two frequencies which correspond to a phase change of π is given by

$$\Delta f = \frac{cd}{4\pi h_E h_R},$$

f is a function of d and h_E or h_R, and these quantities alter as the missile travels.

With phase diversity, the information is retransmitted in a sequential fashion. The interval between two successive transmissions is chosen as a function of the propagation anomalies which one is liable to encounter.

To benefit from the diversity, equipment which is sensitive to the cleanest signal is employed in the system.

Two possibilities are open for this; either the least noisy of the transmission channels is selected by commutation, or a "combiner" used, in which the signals coming from the different transmission channels are combined to provide an improved signal-to-noise ratio.

The order of diversity indicates the number of transmission channels which are simultaneously conveying the information; for reasons of economy it rarely exceeds four.

With a commutation device, the signal-to-noise ratio of the complete transmission is equal to that of the least noisy channel. If, however, the two signals are added according to a quadratic law, it can be shown that even

the more disturbed channel makes a contribution towards improving the signal-to-noise within the diversity-combiner.

The combined signal-to-noise ratio, $(S/N)_{comb}$, with diversity, is expressed in terms of the individual signal-to-noise ratios $(S/N)_i$ of the individual channels by the expression

$$\left(\frac{S}{N}\right)_{comb} = \left(\sum_{1}^{n} (S/N)_i^2\right).$$

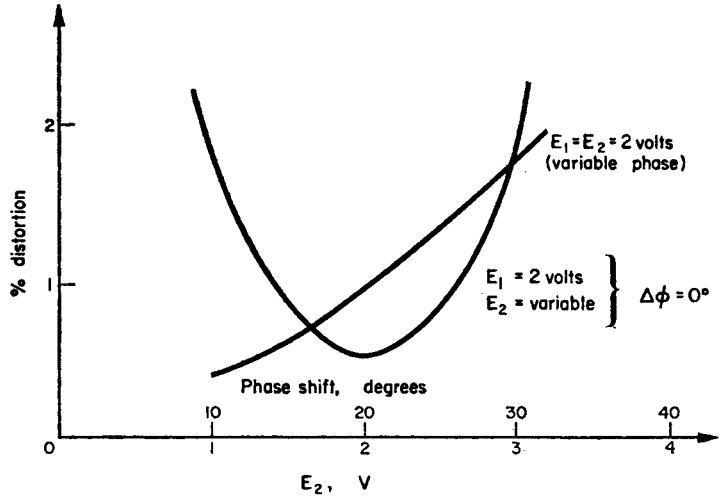


FIG. V.25. Percentage distortion at output of a diversity combiner as a function of the phase difference and the levels of the two signals.

When the signal-to-noise ratios of all the transmission channels are equal, the resulting improvement, as a function of the degree of diversity, is shown in the following table:

Order of diversity	Improvement in the signal-to-noise ratio in db
1	0
2	3
3	4·7
4	6

When frequency modulation is used, the signal S is constant along each channel, and

$$\left(\frac{S}{N}\right)_{comb} = S\sqrt{\left(\sum_{1}^{n} \frac{1}{(N_i)^2}\right)}.$$

If, however, the signals S along the various channels differ widely as a result of fading, the performance of a diversity-combiner tends to show very little improvement over commutation.

It is part of the very principle of diversity that a distortion should appear at the combiner stage if the signals to be combined should be other than exactly in phase and of equal magnitude. In Fig. V.25 are plotted the distortion curves in the case of diversity of the second order.

In the above diagram, the signals E_1 and E_2 at the input of the combiner are sinusoidal, and the phase and amplitude of E_2 are varied with respect to those of E_1, which are regarded as constant.

If supplementary signals are combined with E_1 and E_2, whose values of phase difference and amplitude difference are smaller, the distortions are reduced.

CHAPTER VI

RADIO LINKS FOR REMOTE-CONTROL AND TELEMETRY EQUIPMENT, REGARDED AS MULTIPLEX SYSTEMS

THIS chapter is concerned with the problem of how several messages may be transmitted by the same carrier. These messages are such remote-control commands such as "dive", "climb", "left" and "right"; or they may equally well be measurements, for example of pressure altitude and vibration. The procedure adopted when several channels are grouped together onto a single carrier is known as *multiplexing*; each type of message will be transmitted along a different channel.

We shall distinguish between distribution over frequency and distribution in time.

In this chapter the merits of different kinds of multiplexing are also discussed; in particular in terms of the intermodulation which may occur between different channels.

VI.1. MULTIPLEXING BY DISTRIBUTION OVER FREQUENCY

Each channel or limited family of channels separately modulates a subcarrier wave. The different subcarrier waves are staggered in frequency. The difference between any pair of subcarrier waves must be such as to reduce to a minimum the overlapping of the spectra belonging to adjacent subcarrier waves. The subcarrier waves are mixed forming a complex signal which is in turn used to modulate a carrier wave.

We shall say that the modulation is MX–MY when the subcarrier wave is modulated by the signal in the form MX (which might for example be AM, FM or SSB) while the subcarrier waves modulate the carrier wave in the form MY.

VI.1.1. *Intermodulation Between Channels*

The limitations imposed on information by non-linearity were seen in IV.4, but only for the case of a single signal.

In the most general case, the output signal S is related to the input signal E by the relation

$$S = a_1E + a_2E^2 + a_3E^3 + \cdots + a_nE^n.$$

When the input signal is composed of several transposed frequencies, we have

$$E = e^{j\omega_1 t} + e^{j\omega_2 t} + \cdots + e^{j\omega_p t}$$

and the output signal becomes

$$\begin{aligned} S = {} & a_1[e^{j\omega_1 t} + e^{j\omega_2 t} + \cdots + e^{j\omega_p t}] \\ & + a_2[e^{j\omega_1 t} + e^{j\omega_2 t} + \cdots + e^{j\omega_p t}]^2 \\ & + \cdots \\ & + a_n[e^{j\omega_1 t} + e^{j\omega_2 t} + \cdots + e^{j\omega_p t}]^n. \end{aligned}$$

The amplitude corresponding to the fundamental terms $f_1, f_2, \ldots, f_p$ is a_1; corresponding to the second harmonic $2f_1, 2f_2, \ldots, 2f_p$, it is a_2; and correspondingly for the higher harmonics.

In addition to harmonics of the fundamental frequencies linear combinations of these frequencies of the form

$$\begin{aligned} & f_1 \pm f_2, f_1 \pm f_3, \ldots, f_1 \pm f_p \\ & f_2 \pm f_3, f_2 \pm f_4, \ldots, f_2 \pm f_p \\ & f_1 \pm f_2 \pm f_3, \ldots, f_1 \pm f_2 \pm f_p \\ & f_1 \pm f_3 \pm f_4, \ldots, f_1 \pm f_3 \pm f_p \end{aligned}$$

are introduced by the non-linearity.

Normally, the frequencies of the subcarrier waves are produced by oscillators which are independent of one another, and which thus have arbitrary phases.

If the number of intermodulation products is high, they will all combine together into white noise.

As an example, we consider three frequencies only, and $n = 3$:

$$\begin{aligned} S = {} & a_1(e^{j\omega_1 t} + e^{j\omega_2 t} + e^{j\omega_3 t}) \\ & + a_2(e^{j\omega_1 t} + e^{j\omega_2 t} + e^{j\omega_3 t})^2 \\ & + a_3(e^{j\omega_1 t} + e^{j\omega_2 t} + e^{j\omega_3 t})^3 \\ = {} & a_1 \sum_{i=1}^{3} e^{j\omega_i t} \\ & + a_2 \left[\sum_{i=1}^{3} e^{j2\omega_i t} + 2 \sum_{i,k} e^{j(\omega_i + \omega_k)t} \right] \\ & + a_3 \left[\sum_{i} e^{j3\omega_i t} + 3 \sum_{i,k} e^{j(2\omega_i + \omega_k)t} \right]. \end{aligned}$$

From this example it is seen that the amplitude of the intermodulation product of the form $(2\omega_i + \omega_k)$ is three times that of the corresponding third harmonic.

The amplitudes of the intermodulation combinations are greater than those of their associated harmonics.

When the number of subcarrier waves is reduced, at most only the third harmonic need be considered; the problem at the preliminary design stage is therefore that of selecting the values of the subcarriers to minimize the influence of the intermodulation terms.

VI.1.2. *Selection of Channels for Remote-control and Telemetry Equipment with Several Subcarrier Waves*

It will be assumed that the amplitudes of the signals of which the whole multiplex group is composed are all equal to unity. The composite signal, including all the transposed frequencies $\omega_1, \omega_2, \ldots, \omega_N$, can be written in the form

$$x = \sum_{i=0}^{N} \cos \omega_i t$$

and, writing $x_i = \cos \omega_i t$, this becomes

$$x = \sum_{i=0}^{N} x_i .$$

The signal at the output is therefore expressed by

$$y = \sum_{k=0}^{\infty} a_k x^k .$$

We make use of the basic relationship

$$\cos A \cos B = \tfrac{1}{2}[\cos (A + B) + \cos (A - B)]$$

$$= \tfrac{1}{2} \sum_{+\text{and}-} \cos (A \pm B)$$

in which the symbol $\sum_{+\text{and}-}$ is to be interpreted as a sum over all possible permutations.

This relationship can be generalized to cover cases where several cosine functions appear:

$$U = (\cos A)^a (\cos B)^b \ldots (\cos Z)^z$$

$$= \frac{1}{2^{a+b+\cdots+z-1}} \sum_{+\text{and}-} \cos (A \pm A \pm A \pm \cdots \pm Z \pm Z \pm \cdots).$$

In the bracketed term, there are:

$$a \text{ terms}, A \pm A \pm A \pm \cdots$$

$$b \text{ terms}, \pm B \pm B \pm B \pm \cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$z \text{ terms}, \pm Z \pm Z \pm Z \cdots.$$

The function x^k is composed of the sum of terms of equation (VI.1), multiplied by the coefficients associated with the binomial expansion ($k = a + b + c + \cdots + z$).

It is well known that a function

$$x^i = \left(\sum_i x_i\right)^k$$

can be expressed as

$$x^k = \sum_k \frac{k!}{|a|!|b|!\ldots|z|!} (x_1)^a (x_2)^b \ldots (x_n)^z$$

in which

$$|a| + |b| = \cdots + |z| = k,$$

and as a result

$$y = \sum_k a_k \frac{1}{2^{|a| + |b| + \ldots + |z|}} \cdot \frac{k!}{|a|!\,|b|!\ldots|z|!}$$

$$\times \cos(a\omega_1 + b\omega_2 + \cdots + z\omega_N)\, t.$$

It has been assumed that the terms a_k decrease rapidly when the index takes higher values, and that the amplitudes of the interference products $\cos(a\omega_1 + b\omega_2 + \cdots + z\omega_N)\, t$ are identical to the product.

$$(\cos \omega_1 t)^{|a|} (\cos \omega_2 t)^{|b|} \ldots (\cos \omega_N t)^{|z|}.$$

For intermodulation to occur, an interference product must lie within the bandwidth, with the requirement

$$|a + b + c + \cdots + z| = 1.$$

Interference only appears, therefore, if the products are superimposed onto one of the original signals. Given that the coefficients of the form a_k decrease very rapidly, terms beyond third order need not usually be considered. And indeed, these are in practice the most troublesome. The intermodulation products increase very rapidly with the number of useless frequencies. For a large number of signals, therefore, one is obliged to regard the spectrum of the multiplex system as a uniform one, and the preceding calculation is difficult to apply.

A choice of frequency distribution is required for remote-control and telemetry equipment, in which the number of channels is limited to six. The distribution is regular when the width of the space between two adjacent transposed frequencies is a constant. In this case it should be mentioned that the signals most likely to be affected by interference products lie at the middle of the transposition.

It is obvious that the level of the interference depends upon the amplitudes of the other signals in the band. The third order products combine as the third power of the signal amplitudes.

The resultant frequency of an interference product is

$$\Omega = a\omega_1 + b\omega_2 + \cdots + z\omega_N$$

with

$$a, b, \ldots, z \gtrless 0.$$

We write

$$\omega_1 = \lambda_1\omega_1$$

$$\omega_2 = \lambda_2\omega_1$$

$$\omega_3 = \lambda_3\omega_1$$

$$\omega_N = \lambda_N\omega_1$$

and

$$\frac{\Omega}{\omega_1} = a\lambda_1 + b\lambda_2 + \cdots + z\lambda_N.$$

The following inequalities must be satisfied:

$$\frac{\Omega}{\omega_1} \neq \lambda_1 = 1$$

$$\neq \lambda_2$$

$$\neq \lambda_3$$

$$\neq \lambda_N$$

Ω/ω_1 can be regarded as a point in N-dimensional space; and this point has to be chosen as far as possible from the hyperplanes defined by

$$\frac{\Omega}{\omega_1} = \lambda_1 = 1$$

$$= \lambda_2$$

$$\ldots\ldots$$

$$= \lambda_N.$$

The ideal position for this point is the centre of the hypersphere which has these hyperplanes as tangents, and this determines the values of the ratios $\lambda, \lambda_2, \ldots \lambda_N$.

To make the bandwidth as small as possible, values of λ which correspond to points on the curve of Fig. VI.1 are chosen.

For a few frequencies only (4) a lesser degree of coupling between the channels may be required, and as a result the bandwidth must be increased, and larger values of λ chosen (see curve 2 of Fig. VI.1).

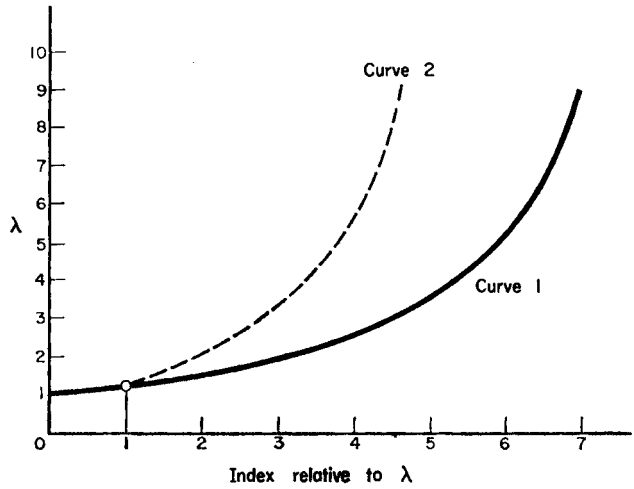


FIG. VI.1. Values of $\lambda_1, \lambda_2, \ldots, \lambda_N$ as a function of the number of transposed frequencies.

In the preceding discussion the problem when it is the frequencies which are modulated has been deliberately omitted. In frequency modulation, the intermodulation product relative to the order k occupies a bandwidth equal to k times the frequency deviation of the subcarriers which are being combined. These, however, are independent of one another, and the spectrum tends to become gaussian when the number of signals which contribute to it becomes large.

VI.1.3. *The Minimum Bandwidth*

When the number of channels is limited to 6, the transposition frequencies, $f_1, f_2, \ldots, f_6$, are determined from the curves shown in Fig. VI.1; the maximum frequency which has to be considered is $f_6 + \mathscr{B}$, where $\mathscr{B}$ is the bandwidth occupied by the modulation. For AM, PM or FM the minimum HF bandwidth is $2(f_6 + \mathscr{B})$.

When the transposition is regular and the multiplex system has n channels, the minimum HF bandwidth to be considered is $\mathscr{B}_{HF} = 2nF_{max}$. In fact, however, the filters employed at the receiver are not perfect, and to compensate for their deficiencies, the transposed frequencies have to be separated by more than twice the maximum LF frequency which is to be transmitted.

We shall adopt a value of $\sqrt{2}$ as coefficient, so that taking into account the supplementary separation which corresponds to a ratio 2,

$$(\mathscr{B}_{HF})_{min} = 4\sqrt{2}nF_{max}.$$

This equation is valid for AM–AM, AM–PM, FM–AM and FM–PM arrangements.

With SSB–AM or SSB–PM multiplexing, the minimum bandwidth is

$$(\mathscr{B}_{HF})_{min} = 2\sqrt{2}nF_{max}.$$

The cross-modulation phenomenon therefore imposes a lower limit on the width of the transmission bands, and further, the modulation and demodulation processes must be linear. It is above all important to avoid over modulation, which is an important source of harmonic distortion. This basically means that an upper limit on the modulation factors of the subcarrier waves is imperative.

When a carrier wave is modulated by a signal which consists of several subcarrier waves, the possible effects can be divided into two classes, depending whether the number of subcarriers is small or great.

When the subcarrier waves are few in number ($n < 10$), the phases of the signals may coincide, and in the limit, the modulation depth for a channel is n times smaller than that of the totality of channels.

When the number of equally transposed channels is great, the probability of all the subcarrier waves being simultaneously in phase at the same time is very small.

We condude, therefore, that the result that the root mean square value of the composite signal increases as the square root of the number of channels. The modulation depth of a channel is therefore $\sqrt{n}$ times smaller.

VI.1.4. *Signal-to-Noise Ratios*

For a single channel, the values of the signal-to-noise ratio, with amplitude modulation and frequency modulation respectively, are

$$\left(\frac{S}{N}\right)_{AM} = \frac{P}{2kF_{max}}$$

and

$$\left(\frac{S}{N}\right)_{\text{FM}} = \frac{P}{2kF_{\max}} \cdot 3m^2,$$

$$m = \frac{\Delta f}{F_{\max}}.$$

For a multiplex link in which the channels are transposed singly, the noise on each is always bounded by the useful frequency interval, $O - F_{\max}$; it is the same therefore as in the case of a single link. Further, if the power of the transmitter is assumed to be unaltered, the useful power available to each channel will be smaller. For a few channels only, therefore, ($n < 10$),

$$\left(\frac{S}{N}\right)_{\text{AM}} = \left(\frac{1}{n}\right)^2 \frac{P}{2kF_{\max}}$$

and for a larger number of channels

$$\left(\frac{S}{N}\right)_{\text{AM}} = \left(\frac{1}{n}\right) \frac{P}{2kF_{\max}}.$$

With frequency modulation, the noise N_k in a channel acts over a band width $f_k - (f_k + F_{\max})$; f_k being the transposition frequency. The channels with higher transposition frequencies are in less favoured positions, since in the following equations, f_k is seen to be in the denominator:

$$\left(\frac{S}{N_k}\right)_{\text{FM}} = \frac{1}{n} \frac{P}{2kF_{\max}} \left(\frac{\Delta F}{f_k}\right)^2 \quad \text{for} \quad n < 10,$$

and

$$\left(\frac{S}{N_k}\right)_{\text{FM}} = \frac{1}{\sqrt{n}} \frac{P}{2kF_{\max}} \left(\frac{\Delta f}{f_k}\right)^2 \quad \text{for} \quad n > 10.$$

The improvement afforded by frequency modulation over amplitude modulation is defined by

$$G_{\text{FM/AM}} = \frac{(S/N)_{\text{FM}}}{(S/N)_{\text{AM}}} = m^2 = \left(\frac{\Delta f}{f_k}\right)^2.$$

More generally, the improvement in signal-to-noise for an *MX–MY* arrangement can be written

$$G_{MX-MY/\text{AM}} = \frac{(S/N)_{MX-MY}}{(S/N)_{\text{AM}}} = K_1 M_1 K_2 M_2$$

in which K_1 and K_2 are constants associated with the first and second kinds of modulation, and M_1 and M_2 are the corresponding modulation indices.

The values of G, due to B.D.Holbrook and J.T.Dixon, are summarized in the following table.

TABLE VI.1.

	Few channels $n < 10$	Many channels $n > 10$
$G_{AM/AM}$	n^{-2}	n^{-1}
$G_{FM/AM}$	$0{\cdot}75\ F_{max}^{-2}\mathscr{B}^2 n^{-2}$	$0{\cdot}75\ F_{max}^{2}\mathscr{B}^2 n^{-1}$
$G_{AM-AM/AM}$	$0{\cdot}125\ n^{-2}$	$0{\cdot}062\ n^{-1}$
$G_{AM-PM/AM}$	$0{\cdot}031\ n^{-2}r^2y^2$	$0{\cdot}011\ n^{-1}r^2y^2$
$G_{FM-AM/AM}$	$0{\cdot}093\ n^{-2}r^2$	$0{\cdot}011\ n^{-1}r^2$
$G_{FM-PM/AM}$	$0{\cdot}375\ n^{-2}r^2$	$0{\cdot}035\ n^{-1}r^2$
$G_{SS-AM/AM}$	$x(n)\cdot n^{-1}$	$x(n)\ n^{-1}$
$G_{SS-PM/AM}$	$0{\cdot}75\ x(n)\ n^{-1}r^2y^2$	$0{\cdot}75\ x(n)\ n^{-1}r^2y^2$

In this table, n is the number of channels, F_{max} is the maximum LF frequency transmitted by a channel, y is the safety factor which we have taken equal to $\sqrt{2}$, $\mathscr{B}$ is the IF bandwidth and $r = \mathscr{B}n^{-1}F_{max}^{-1}$. The values of $x(n)$ are as follows:

n	1	4	12	48	72	∞
$x(n)$	1	1·6	2·5	5	6·2	16

VI.2. MULTIPLEX SYSTEMS WITH TEMPORAL DISTRIBUTION

In the preceding section, it was seen that a message can be transmitted by sampling the input quantity by means of pulses, subject to the condition that the sampling frequency must be equal to at least twice the maximum frequency, F_{max}, of the input signal. We denote the sampling or repetition frequency by f_r, and choose it to be

$$f_r = 3F_{max}.$$

Between each sampled pulse there is an empty space, which can be used for the interpolation of other message pulses, or in other words, of other channels. In this way we obtain a multiplex system with the various channels distributed in time (see Fig. VI.2).

Each channel is sampled at the same frequency, f_r and the space between two pulses is characterized by T_r, the interval between two adjacent pulses.

If the channels are to be easily distinguishable, it is necessary to provide a synchronization pulse, distinct from the others; the various channels are then distinguished by measuring their phases with respect to that of the

synchronization pulse. The latter is recognizable from among the channel pulses by its shape (it may be longer or greater in amplitude).

When the multiplex has n channels, therefore, the repetition frequency of the pulses is given by

$$f_r = 3nF_{max}.$$

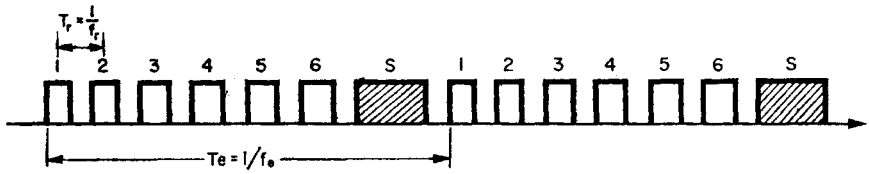


FIG. VI.2. Multiplexing with time distribution.

This relationship is valid for PAM, PPM, PFM and PDM multiplex equipment. PNM (pulse number modulation) systems and PCM (pulse code modulation) systems require m pulses to represent 2^m distinct levels, and hence for n different channels, the repetition frequency becomes

$$f_r = 3mnF_{max}.$$

VI.2.1. *Cross-modulation*[33]

Cross-modulation in temporally distributed multiplex systems arises from residual signals from the adjacent channels. These parasitic signals, whose form is exponential, interfere with the required signal as shown in Fig. VI.3.

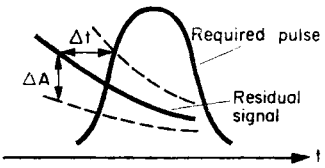


FIG. VI.3. Perturbation of wanted pulse by residual signal.

It can be grouped into three main categories: low frequency trailing and a group arising from various other causes.

VI.2.1.1. *Low Frequency Cross-modulation*

This arises from the RC networks in video-amplifiers. Figure VI.4 shows the way in which an RC circuit affects the input pulse.

During the interval θ_1, the grid potential obeys a law of the form

$$v_g = k\,e^{-t/RC} = (a - A)\,e^{-t/RC},$$

so that for $t = \theta_1$,

$$v_g = (A - a)\,e^{-\theta_1/RC},$$

and during the interval θ_2,

$$v_g = [(a - A)\,e^{-\theta_1/RC} + A]\,e^{-t/RC}.$$

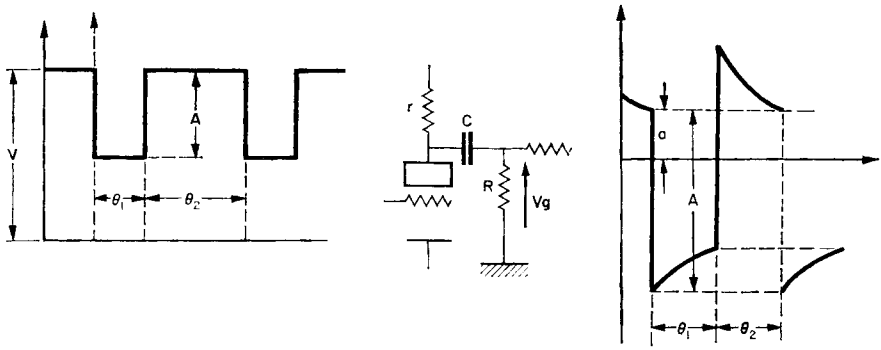


FIG. VI.4. Waveform distortion in RC circuits.

At time $t = \theta_2$, the initial potential should be $v_g = a$ again. This gives

$$a = A\,\frac{e^{-\theta_2/RC} - e^{-(\theta_1+\theta_2)/RC}}{1 - e^{-(\theta_1+\theta_2)/RC}}.$$

In fact, the time constant RC is always much greater than either θ_1 and θ_2, so that

$$a = A\,\frac{\theta_1}{\theta_1 + \theta_2}.$$

Suppose that one of two adjacent pulses undergoes an instantaneous variation of amplitude A; accordingly the channel is perturbed in the following way:

$$\Delta V_g = \Delta A[1 - e^{-\theta_1/RC}]\,e^{-\theta_2/RC}$$

and by neglecting all but the first terms of the series expansions for the exponentials, this reduces to

$$\Delta V_g = \Delta A\,\frac{\theta_1}{RC}.$$

The cross-modulation is given by

$$\frac{\Delta V_g}{\Delta A} = \frac{\theta_1}{RC}$$

and so long as θ_1/RC is constant, the cross-modulation is linear.

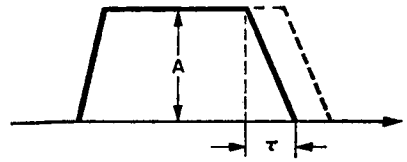


FIG. VI.5. Modulation of the falling edge duration of a pulse.

When n is the number of channels and when these are not modulated at the same frequency and have arbitrary phases, the powers combine quadratically. This is equivalent to increasing the cross-modulation itself by $\sqrt{n}$. If the amplifier contains N stages, the total cross-modulation per channel will be

$$(D_{max})_{PAM} = N\sqrt{n}\,\frac{\theta_1}{RC}.$$

Consider now the situation when the modulation in the channels is of the pulse length. This modulation might, for example, be applied to the falling edge of the pulse (see Fig. VI.5).

As before

$$v_g = (a - A)\,e^{-(\theta_1+\theta_2)/RC} + A\,e^{-\theta_2/RC}.$$

Suppose that θ_1 increases by an amount $\Delta\theta_1$, $(\theta_1 + \theta_2)$ remaining constant. Then

$$\Delta V_g = -\frac{A}{RC}\,e^{-\theta_2/RC}\Delta\theta_2 = \frac{A}{RC}\,e^{-\theta_2/RC}\Delta\theta_1,$$

or

$$\Delta V_g \simeq A\,\frac{\Delta\theta_1}{RC}.$$

This amplitude variation results in a phase change $\Delta\theta_1'$ at the face of the adjacent pulse, such that

$$\frac{\Delta V_g}{\Delta\theta_1'} = \frac{A}{\tau}$$

or

$$\frac{\Delta\theta_1'}{\Delta\theta_1} = \frac{\tau}{RC};$$

in which τ is the rise-time of the pulse; for N stages and n channels, therefore, the maximum cross-modulation per channel is given by

$$(D_{max})_{PDM} = N\sqrt{n}\,\frac{\tau}{RC}.$$

The cross-modulation coefficient can be reduced by shortening the rise-time τ of the pulse.

With position modulation, it can be shown in a similar way that the cross-modulation coefficient is

$$(D_{max})_{PDM} = N\sqrt{n}\,\frac{\tau}{RC}\cdot\frac{\theta_1}{RC}.$$

VI.2.1.2. *Trailing Cross-modulation*

The spectrum of the pulses is more widely spread when the rise-time is shorter. Since the bandwidth of an amplifier is limited, the components of the spectrum outside the band must be attenuated, and the pulse, therefore, distorted. This distortion has the appearance of a slow tail at one end, (see Fig. VI.6) and this encroaches upon the zone occupied by the next channel.

The time constants, γ_r are here much shorter than in the case of LF cross-modulation.

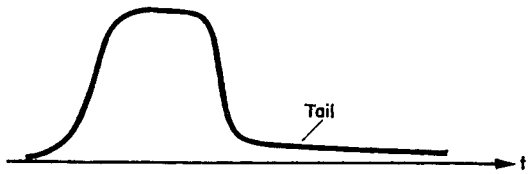


FIG. VI.6. Pulse tail produced by limited amplifier bandwidth.

The response of an amplifier to a pulse is of the form

$$F(t) = A(1 - e^{-t/\gamma_r}) \qquad \text{for the rising edge.}$$

$$F(t) = A(1 - e^{-(t-\theta_1)/\gamma_r}) \qquad \text{for the falling edge.}$$

The rise-time τ must be considerably shorter than the pulse length θ_1.

The residual signal in amplitude modulation is effectively given by

$$\left(\frac{\Delta v}{\Delta A}\right)_{av} = \frac{\gamma_r}{\theta_1}\,e^{-\theta_2/\gamma_r}.$$

For N stages, with n-channel multiplex, the cross-modulation coefficient per channel is

$$(D_{max})_{PAM} = N\frac{\gamma_r}{\theta_1}\,e^{-\theta_2/\gamma_r}$$

and is thus independent of the number of channels.

With pulse duration or pulse position modulation, we find

$$(D_{max})_{PDM \text{ and } PPM} = N \frac{\tau}{\gamma_r} e^{-\theta_2/\gamma_r}.$$

Since θ_2 varies with the modulation, the above formula gives the maximum distortion when θ_2 is a minimum. In the remaining cases, the distortion ceases to be linear.

VI.2.1.3. *Various Other Types of Cross-modulation*

The instruments used as "modulation converters" are generally not linear, and introduce intermodulation noise.

An "amplitude-duration" converter converts amplitude modulation into pulse duration modulation; between one pulse and its successor, a residual signal remains which is proportional to the maximum amplitude of the pulse, and this affects the rising edge of the following pulse—the duration of the transformed pulse is, in consequence, also affected. This residue constitutes the cross-modulation of the converter.

Where "duration-position" converters are concerned, the residue of the differentiated pulse is of the form

$$u = \frac{\gamma_r}{\tau} e^{-t/\gamma_r},$$

so that at the end of a time $\theta = \theta_1 + \theta_2$,

$$u = \frac{\gamma_r}{\tau} e^{-\theta/\gamma_r}.$$

When there is an increase $\Delta\theta$, due to the modulation of a perturbing channel, the amplitude of the residue will be

$$\Delta u = -\frac{1}{\tau} e^{-\theta/\gamma_r} \Delta\theta,$$

and this variation Δu results in a shift of the edge by an amount

$$\frac{\Delta u}{\Delta\theta_1} = \frac{du}{dt} = \frac{1}{\tau} e^{-t/\gamma_r}.$$

VI.2.1.4. *The Consequences of Intermodulation; Choice of Pulse Width*

If the pulses were perfectly rectangular, the maximum width which an unmodulated pulse could be given would be $T_r = 1/f_r$. To prevent the channels from encroaching upon one another, the instants at which two successive unmodulated pulses appeared would have to be separated by T_r. The pulses

are not, in fact, rectangular; and furthermore, the modulation itself shifts the leading edges of the pulses in pulse position and pulse duration modulation, and this increases the possibilities of overlapping. A safety margin must therefore be provided, and one is obliged to shorten the pulse-length to a value $T_i < T_r$.

We arrange that $T_i \leqslant kT_r$, in which k is a coefficient smaller than unity.

The maximum values of k for which the cross-modulation separation is better than 60 db have been calculated by Landon, and are set out in Table VI.2.

TABLE VI.2.

Modulation	k_{max}
PAM–AM	0·25
PAM–FM	0·5
PDM–AM	0·125
PDM–FM	0·125
PPM–AM	0·097
PPM–FM	0·097
PNM–MX	0·333
PAM–PCM–MX	0·333

With the aid of the values given in this table, T_i can be deduced once $T_r = 1/3nF_{max}$ is known. The minimum HF bandwidth is then $(\mathscr{B}_{HF})_{min} = 1/T_i$.

VI.2.2. *The Signal-to-Noise Ratio*

The pulse input to the receiver is contaminated with noise, and the waveform is similar in essentials to the waveform drawn in Fig. VI.7. At the receiver, therefore, the pulse must be cleaned up, or shaped, as much as possible; to do this, two methods are available: gating and "slicing" (shaping by clipping and limiting).

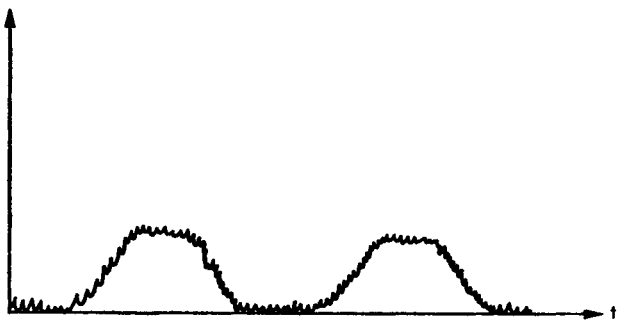


FIG. VI.7. The waveform of a received pulse.

It can be arranged that the receiver remains blocked in the absence of a pulse; or even better, in the case of PAM modulation, the pulse which has been received can be gated by a reading pulse, which is narrower, and which eliminates the noise on the edges (see Fig. VI.8). It is clearly not possible to do this when the modulation affects the edges of the pulses.

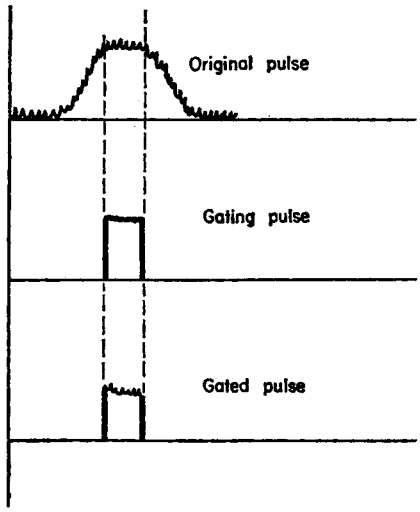


FIG. VI.8. Elimination of noise by gating for PAM pulses.

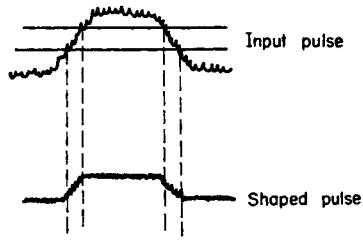


FIG. VI.9. Half-height slicing.

When the modulation is PFM, PDM or PPM, the pulses can be sliced at about half their height, as shown in Fig. VI.9. By doing this, the background noise is suppressed between the pulses and during the central part of the pulse, when it has its peak value. The noise then only affects the edges of the truncated pulse, as can be seen in the diagram.

In multiplex equipment using coded modulation (PCM or PNM), the desired information is expressed by the presence or absence of a pulse. This kind of yes–no transmission allows the simultaneous use of limiting, gating

and slicing. Ideally it is therefore possible to produce perfect pulses, free of noise. If, in fact, the amplitude of a noise pulse is equal to half the useful level, and is of opposite sign, the code pulse will, after all, be affected by the noise. There is, therefore, clearly a limit to the signal-to-noise ratio.

Despite these efforts to reduce noise, the signal-to-noise ratio always has an appreciable value, which is more or less large according to the type of modulation selected. If the above improvements are taken into account, and if the bandwidth is given the value calculated in the preceding section, the improvements in signal-to-noise ratio, compared with amplitude modulation, are those summarized in the following table.

TABLE VI.3

Transmission system	$n > 10$
G PAM-AM/AM	$1/n$
G PAM-FM/AM	$0{\cdot}068\, n^{-1}r^2$
G PPM-AM/AM	$0{\cdot}048\, n^{-1}r^2$
G PPM-FM/AM	$0{\cdot}435\, n^{-1}r$
G PNM-XX/AM	∞
G PAM-PCM-XX/AM	∞
G PDM-AM/AM	$0{\cdot}168\, n^{-1}r$
G PDM-FM/AM	$3{\cdot}23\, n^{-1}r$

n: number of channels
F_{max}: maximum frequency transmitted by the channel
$\mathscr{B}_{HF}$: IF bandwidth
$r = \mathscr{B} n^{-1} F_{max}^{-1}$.

This indicates that in coded modulation, the signal-to-noise ratio is infinite. In practice the quantization noise must be taken into account.

The modulating signal is of the form:

$$x = A \sin \Omega t.$$

After quantization, it becomes

$$y = x + \varepsilon(x)$$

in which $\varepsilon(x)$ represents the error introduced by quantization,

$$\varepsilon(x) = \frac{\alpha}{2} \sum_{n=1}^{\infty} (-1)^n \frac{\alpha}{n\pi} \sin \frac{2n\pi}{\alpha} x.$$

After decomposing the amplitude A into p quantization levels, we find $A = p\alpha$.

It can be shown that the minimum value, K, of the ratio of the signal amplitude, $p\alpha$, to that of a harmonic is

$$K \simeq 9p^{3/2}.$$

Further, the total energy within the error $\varepsilon(t)$ can be shown to have the value

$$\overline{\varepsilon(t)^2} = \frac{1}{T}\int_0^T \varepsilon^2(t)\,\mathrm{d}t \simeq \frac{\alpha^2}{12}.$$

Since the signal energy, S, is

$$S = \tfrac{1}{2}p^2\alpha^2,$$

the signal-to-noise ratio will be

$$(S/N)_{\min} = 6p^2,$$

and if m is the number of lineary digits in the code,

$$p = 2^{(m-1)}$$

and hence

$$(S/N)_{\min} = (6m + 2)\ \mathrm{db}.$$

Note: In systems in which the pulses are sliced at half their height, it is the amplitude of the signal at the shaping level which has to be considered when evaluating the signal-to-noise ratio.

VI.2.3. *Sampling and Distribution Procedures*

One method of pulse multiplying is to apply the basic pulse, of repetition frequency f, to a delay line with n tappings, (n being the number of channels). The delays are $\theta, 2\theta, 3\theta, \ldots, n\theta$; and the resulting n pulses are modulated by the n LF inputs signals.

By adding the signals leaving the modulators, we obtain multiplexing in the form of a sequence of pulses (see Fig. VI.10).

The generator can consist of a suitably driven bistable multivibrator, (binary). A transistor binary circuit is shown in Fig. VI.11. When the transistor $T1$ is cut-off, $T2$ conducts and vice versa. Cut-off of the transistor $T1$ is ensured by the collector potential produced by the current from $T2$ passing through R_E. If the multivibrator is to work properly, the following inequality must be satisfied:

$$R_c C > \frac{1}{f_\alpha}$$

in which f is the cut-off frequency of the transistor.

The time constant, $R_E C_E$, must be at least equal to the commutation time.

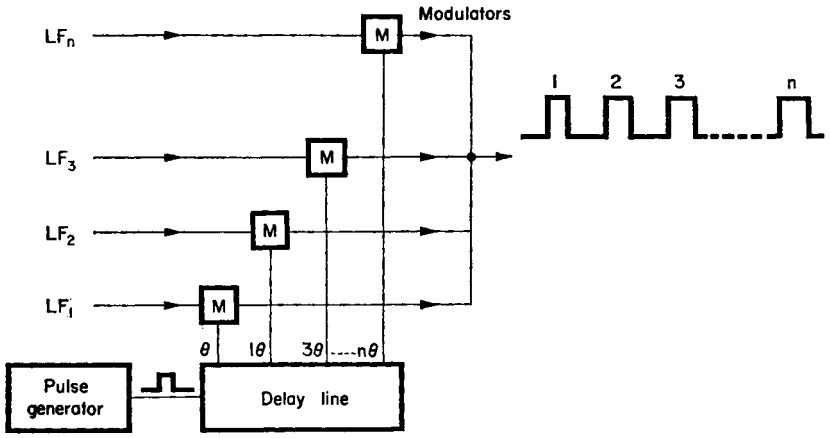


FIG. VI.10. Multiplexing with time distribution.

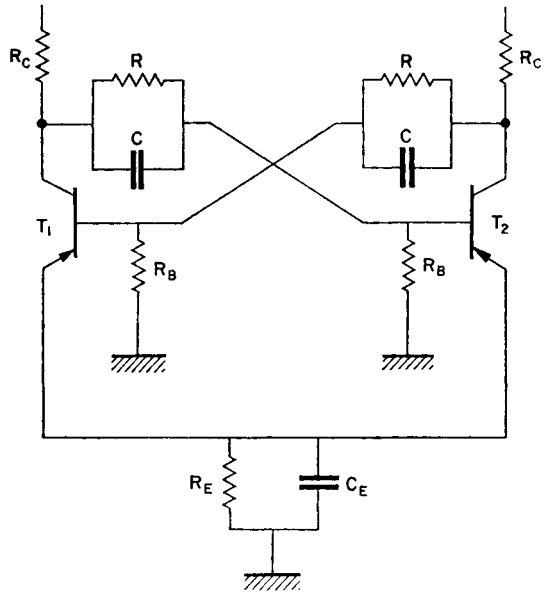


FIG. VI.11. A transistorized binary.

The binary is driven by a periodic voltage; alternatively the combination can be replaced by an astable multivibrator (free-running).

The delay line usually consists of a sequence of lumped components with constant values of L, M, C, as shown in Fig. VI.12.

A number of results have been given by Libois for the design of delay lines.

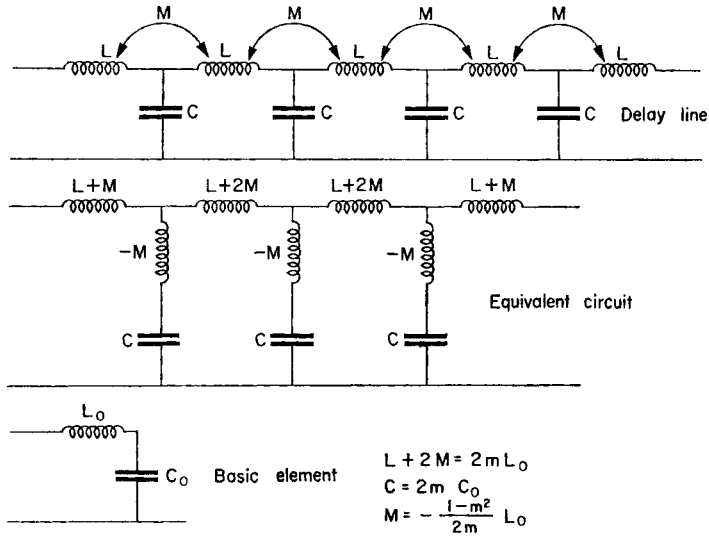


FIG. VI.12. Delay line with lumped parameters.[53]

For the value $m = 1{\cdot}27$ we find

– the mutual inductance between two adjacent self-inductances,

$$\frac{M}{L} = 0{\cdot}12$$

– the cut-off frequency,

$$\omega_c = 1{\cdot}15 \frac{2}{\sqrt{(LC)}}$$

– the phase delay per basic element,

$$\tau = \frac{m}{\omega_c} = 1{\cdot}1 \sqrt{(LC)}$$

– the characteristic impedance of the line,

$$Z_c = \sqrt{\left(\frac{L_0}{C_0}\right)} = 1{\cdot}1 \sqrt{\left(\frac{L}{C}\right)}.$$

If r is the resistance of the inductor L, and if x is the ratio of the working frequency and the cut-off frequency of the line, f_c ($x = f/f_c$), the attenuation when x is very much less than unity is given by

$$A = \frac{mr}{2Z_c}(1 - x^2)^{-1/2}.$$

Modulators vary according to the type of modulation which is employed. For further information, the reader is referred to the chapter on the different types of modulation.

VI.2.4. *Synchronization*

The use of a synchronization pulse is imperative when pulses from different channels have to be distinguished from one another at the receiver. To be able to pick out this reference pulse from amongst the rest, it can be given a different shape; or several pulses, resembling those used in the multiplex channels, can be grouped in a special way.

The synchronization signal may be distinguished by having a greater amplitude than the remainder, but this is not generally desirable since the pulses will be shaped at the receiver to reduce the effect of noise.

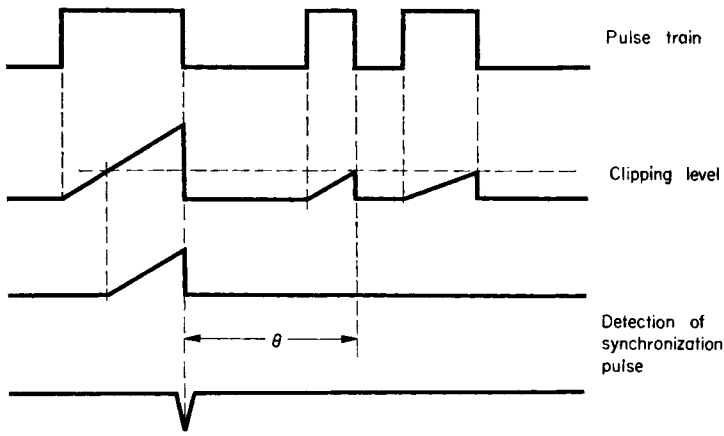


FIG. VI.13. Detection of synchronization pulse.

On the other hand it is easy to detect a pulse which persists longer than the others. At the receiver it is only necessary to change the pulses into saw-tooth form (see Fig. VI.13) for the longer pulse to be identifiable by its amplitude.

To separate the various channels, the time interval separating the falling edge of each channel pulse from the synchronization pulse has to be noted.

The synchronization "message" can alternatively be sent as a pair of pulses, identical with those in the channels, but fixed at a constant distance apart.

Let θ_1 be the length of each pulse and θ_2 the distance between them ($\theta = \theta_1 + \theta_2$). The selection at the receiver is made by superimposing the synchronization signal onto an identical pulse, a distance θ away (see Fig. VI.14).

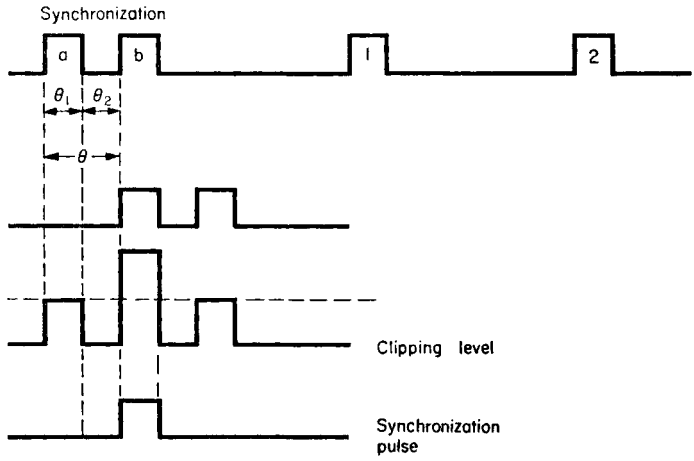


FIG. VI.14. Detection of a synchronization pulse consisting of a pair of pulses.

The pulses in the channel are classed with respect to the pulse (*b*) of the synchronization signal at the transmitter. To avoid any risk of the shift θ producing interference between the displaced synchronization pulse (*b*) and channel (1) it is necessary to ensure that $\theta < T_g$.

T_g represents a safety time between the channel pulses.

VI.3. THE THRESHOLD RANGE OF A RADIO LINK

The problem is to calculate the maximum distance or range which can be reached, when the power and gain at the transmitter are P_E and G_E respectively, taking into account the minimum power which is acceptable at the receiver for a given signal-to-noise ratio. This distance will be an idealized value, as in practice the type of propagation would have to be taken into account and also the effect of fading of the transmission, although the latter is difficult to include.

VI.3.1. *Minimum Power at the Receiver to Ensure Adequate Protection against Noise*

The signal-to-noise improvement with frequency modulation, $G_{FM/AM} = 3\ m^2$ is only valid when the amplitude of the carrier wave is considerably larger than the noise level. For example, frequency modulation becomes useful when the (S/N) input $\geqslant 9$ db. Similarly, for pulse modulation, it has been seen that slicing at about half amplitude is effective only if the noise peaks are lower than this level. Further, when noise peaks reach the amplitude of the signal pulses, the synchronization devices can no longer work. We shall subsequently assume, therefore, that the signal-to-noise ratio is greater than 9 db. The relationships of Tables VI.1 and VI.3 are valid only under these conditions.

The peak power of a sinusoidal carrier wave is equal to twice the mean power. Where background noise is concerned, the peak power is sixteen times greater, under the above conditions, than the root mean square noise power. In periods of intense noise even this value can be exceeded, but with a probability of less than one part in 10^4.

The background noise at the input of a receiver is equal to

$$P_N = FkT\mathscr{B}$$

in which F is the noise factor, T the absolute temperature, $k = 1{\cdot}381 \times 10^{-23}$ joules/degree and $\mathscr{B}$ is the frequency bandwidth of the receiver.

For frequencies higher than 500 Mc/s, the noise factor exceeds 10 db. We shall take F arbitrarily equal to 16. This value covers most cases. The minimum power at the receiver is therefore

$$(P_R)_{\min} = 8P_N,$$
$$= 3{\cdot}2 \times 10^{-20}\, F\mathscr{B} \text{ watts}$$

when $\mathscr{B}$ is expressed in cycles per second.

It should be noted that for pulse modulation, the transmitter is inoperative except during the emission of a pulse. The mean power, therefore is reduced in the ratio $\mu = f_{r/\mathscr{B}}$. When the receiver input is sliced at half amplitude, the signal power must be four times as great. As a result

$$(P_R)_{\min} = 3{\cdot}2 \times 10^{-20}\, \varrho F\mathscr{B}$$

in which $\varrho = \mu$ for the PAM–AM system
$\varrho = 4\mu$ for the PPM–AM, PFM–AM, PDM–AM, PNM–AM and PAM–PCM–AM systems
$\varrho = 1$ for the other systems.

VI.3.2. *Minimum Power at the Receiver to ensure a Given Signal-to-Noise Ratio in the LF Channels*

It is desirable to establish a given LF signal-to-noise ratio in the n channels of the multiplex systems to allow the corresponding HF bandwidth to be calculated.

An amplitude modulated HF link has two sidebands, $f_0 \pm F_{max}$. The noise in the two bands adds quadratically, so that the noise is $\sqrt{2}$ times as large as in a single sideband. In other words the LF signal-to-noise ratio q_0 is $(2/\sqrt{2})^2 = 2$ times higher than the HF sidebands-to-noise ratio:

$$q_0 = 2\left(\frac{\text{sidebands}}{\text{noise}}\right)_{HF} = \left(\frac{\text{carrier}}{\text{noise}}\right)_{HF} = \frac{P_R}{2FkTF_{max}}.$$

In the following, the LF bandwidth is regarded as being twice the maximum LF frequency.

The power necessary to ensure a signal-to-noise ratio of q_0 with the AM arrangement is therefore taken as a standard of reference

$$P_R = 2FkTF_{max}q_0.$$

Denoting the HF signal-to-noise ratio improvement of the instrument in question by G, the mean power necessary at the receiver is therefore

$$P_R = 2FkTF_{max}\frac{q}{G}.$$

We have to find the value of the bandwidth $\mathscr{B}_{HF}$ for which P_R has a minimum, knowing that the HF bandwidth must be greater than $(\mathscr{B}_{HF})_{min} = 1/T_i$ to eliminate distortion and intermodulation ($\mathscr{B}_{HF} > (\mathscr{B}_{HF})_{min}$).

The bandwidth has also an upper limit, introduced by practical considerations (presence of IF with wide bands).

$$(\mathscr{B}_{HF})_{max} < 50 \text{ Mc/s}.$$

$$\mathscr{B}_{HF} < (\mathscr{B}_{HF})_{max}.$$

There is yet a third condition to satisfy, that

$$P_R \geqslant (P_R)_{min}$$

with

$$(P_R)_{min} = 3{\cdot}2 \times 10^{-20}\,\varrho F\mathscr{B}_{HF}.$$

In Fig. VI.15 are plotted curves which represent these equalities and inequalities with P_R as ordinate and $\mathscr{B}_{HF}$ as abscissa. The shaded regions are thus forbidden. The point A represents the origin of the working region.

The gain G is that of Tables VI.1 and VI.3. It falls with the number of channels; thus as the number of channels increases, the curve $P_R = f(G)$ shifts upwards. Generally speaking, we always choose the lowest point on the curve $P_R = f(G)$, provided it does not lie in the shaded zone.

If the curve $P_R = f(G)$ is a horizontal straight line (this is the case with AM–AM and PAM–AM equipment), the point which corresponds to the minimum bandwidth is chosen.

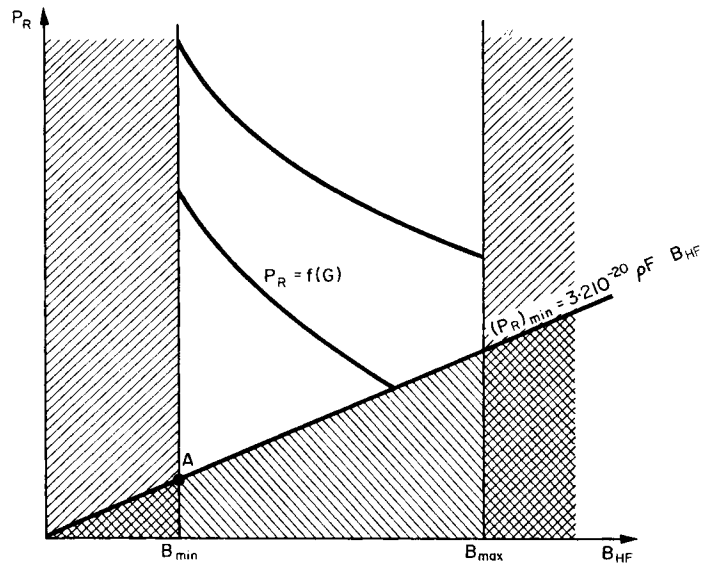


FIG. VI.15. Limits on received power P_R as a function of HF bandwidth.[103]

When, on the other hand the possibility of jamming arises, the narrowest HF band possible is employed, and the value of the power at the receiver will correspond to the point of intersection of the function $P_R = f(G)$ with the vertical $\mathscr{B}_{HF} = (\mathscr{B}_{HF})_{min}$.

J. P. Voge has published a table of experimental results obtained with this method which is reproduced below (p. 168).

VI.3.3. *The Maximum Range of the Link*

The relation between the power at the transmitter and that at the receiver as a function of the distance d, the wavelength λ and the effective areas S_E and S_R of the aerials, is

$$P_E = P_R \frac{\lambda^2 d^2}{S_E S_R}.$$

TABLE VI.4. MINIMUM POWER AT THE RECEIVER in pW (10^{-12}W)

System	$n = 4$	$n = 12$	$n = 48$	$n = 72$
PCM–FM	0·211	0·94	5·12	11·5
PNM–FM				
PAM–PCM–FM	0·46	1·36	5·42	
PNM–AM				
PAM–PCM–AM	0·60	1·8	7·4	
SS–PM	0·43	2·77	515	1760
PAM–FM	0·77	4·97	3050	10,250
PDM–FM	1·12	22	3500	5870
PPM–AM	0·086	6·2	3950	13,500
FM–PM	0·68	8·42	5450	18,300
SS–PM–FM	1·27	20	13,000	44,100
FM–AM	1·08	26	16,600	56,300
PPM–FM	3·03	163	26,000	58,500
AM–PM	1·97	53	34,000	116,000
PDM–AM	7.1	615	98,600	222,000
SS–AM	574	17,200	690,000	1,040,000
PAM–AM	574	17,200	690,000	1,040,000
AM–AM	18,300	275,000	11,000,000	16,500,000

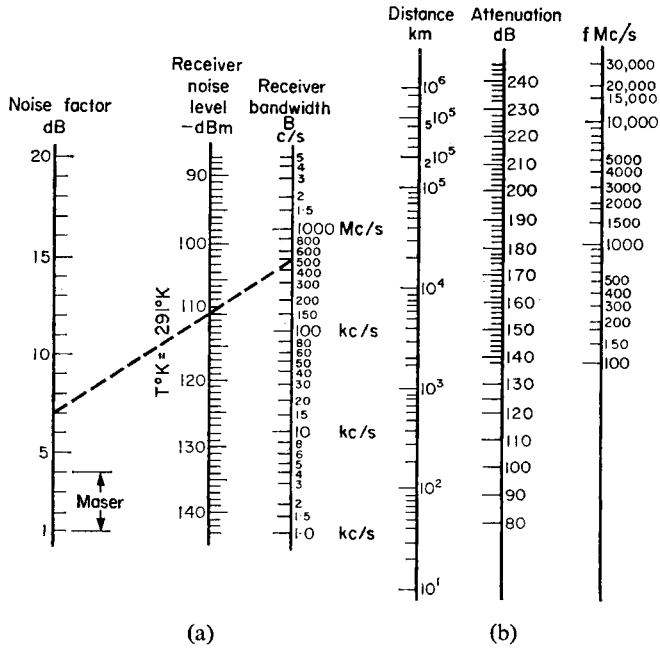


FIG. VI.16. (a) Determination of receiver noise level. (b) Determination of free space attenuation between two isotropic aerials.[79]

Introducing the gains of the aerials, $G_E = \frac{4\pi S_E}{\lambda^2}$ and $G_R = \frac{4\pi S_R}{\lambda^2}$, we have

$$P_E = P_R \frac{(4\pi)^2}{G_E G_R} \left(\frac{d}{\lambda}\right)^2.$$

Usually, two cases are possible at the aerial; either the transmitting and the receiving aerials are both isotropic (omnidirectional) or one of them is directive.

If the diagrams of both aerials are isotropic, we find

$$G_E = G_R \simeq 1$$

thus

$$P_E = P_R \left(\frac{4\pi}{\lambda}\right)^2 d^2.$$

It is seen from the preceding equation that the power required at the transmitter is greater at shorter wavelengths and hence higher frequencies. It would seem advantageous, therefore, to employ low frequency links. However, the dimensions of the missile and of the aerials are limited, and as a result the working frequency may not fall below a certain value. A second limitation, which creates a minimum for the carrier frequency, results from the quantity of information which the carrier has to be able to transmit without distortion. Further, the kind of link to be employed must be taken into account (extra-terrestrial, tropospheric, ionospheric, etc.), as this dictates the frequency range which can be used. A value for the minimum frequency which is compatible with all these conditions has to be deduced.

Two nomographic charts which enable the parameters for a given link to be established readily are given in Fig. VI.16.

Note: The equation for attenuation in free space ignores possible fading, it is advisable to allow an additional margin of safety of 20 db in the power.

VI.4. COMPARISON OF DIFFERENT TYPES OF MODULATION AND MULTIPLEXING

Different types of modulation and multiplexing can be described in terms of a number of different characteristics, such as

- the maximum range
- the bandwidth required by the information
- protection against interference
- the quality of the transmission
- the ease with which it can be achieved.

The relative importance of each of these characteristics depends upon the application. The problems encountered in telemetry and remote-control are thus of different natures.

Protection against deliberate jamming is more likely to be imperative for remote-control than for telemetry. The quantity of information, on the other hand, is usually greater in telemetry than in remote-control, since a greater bandwidth is necessary for the measurement of some phenomena than for the transmission of a remote-control command.

As can be seen from Table VI.4, the minimum power at the receiver is smaller for PPM–AM, PCM–FM, SS–PM, PNM–FM and FM–PM equipment when the number of channels is low ($n = 4$). If the number of channels increases, the minimum power with a PPM–AM link becomes prohibitive, whereas a PNM–AM multiplex is still practicable. It should, however, be noticed that a PCM–FM system, with conversion in time, is, in every case, advantageous, as also are PNM–FM and PAM–PCM–FM.

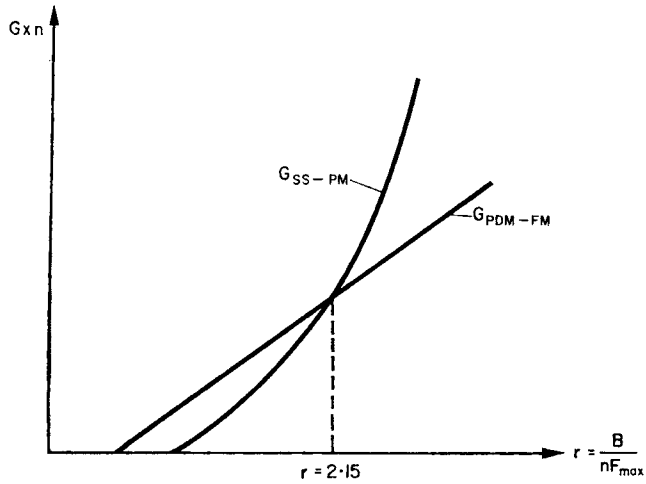


FIG. VI.17. Variation of signal-to-noise ratio improvement as a function of $r = \mathscr{B}/nF_{max}$.

In comparing the improvements in signal-to-noise ratio, three cases are to be distinguished: those in which the gains are of the forms, firstly, $(A/n)r^2$, secondly, $(B/n)r$, and, finally, the cases of PNM–MX and PAM–PCM–MX modulation.

Among modulation devices with a gain of the form $(A/n)r^2$, the highest value for A is 1·5, and corresponds to SS–PM modulation (this is for the single channel case). For systems with a gain of the form $(B/n)r$, the largest value of the coefficient B is associated with PDM–FM for which it is equal to 3·23.

It seems, therefore, that the arrangements to be compared are the multiplex systems SS–PM and PDM–FM. The two gains are equal when the relation

$$G_{SS-PM} = G_{PDM-FM}$$

is satisfied, which implies

$$1{\cdot}50\,\frac{1}{n}\,r^2 = 3{\cdot}23\,\frac{1}{n}\,r,$$

or

$$r = 2{\cdot}15.$$

So long as r remains smaller than 2·15, $G_{\text{PDM–FM}}$ is greater than the gain with SS–PM modulation; for $r > 2{\cdot}15$ the converse is true.

Let us now try to compare pulse duration modulation and coded modulation. It was seen in the chapter on noise that

$$[(S/N)_{\text{LF}}]_{\text{PDM}} = [(S/N)_{\text{HF}}]_{\text{PDM}} \left(\frac{\Delta T_{\text{effective}}}{\varepsilon}\right)^2$$

in which ΔT_{eff} is the effective value of the displacement of a pulse and $\varepsilon = 1/2\mathscr{B}$ is the pulse rise time.

Consider a concrete example. For a twelve channel multiplex, ΔT_{eff} is 2μs, and $\varepsilon = 0{\cdot}2\,\mu$s, hence

$$\left[\frac{(S/N)_{\text{LF}}}{(S/N)_{\text{HF}}}\right]_{\text{PDM}} = 20\text{ db.}$$

for a bandwidth $\mathscr{B} = 2$ Mc/s.

With coded modulation, at least six binary digits are required in the code to ensure that the quality of the link is acceptable. If the synchronization pulse resembles one of the channel pulses,

$$f_r = 6(12 + 1)F_{\text{max}}$$

with $F_{\text{max}} = 10$ kc/s, $f_r = 850$ kc/s. The band occupied by the pulses has a width $\mathscr{B} = 2f_r = 1{\cdot}7$ Mc/s, which is virtually the same as the width we should select with PDM modulation.

If, therefore, it is supposed that the signal-to-noise ratios at the input are equivalent for the two types of modulation, and equal for example to 20 db, it can be seen that the signal-to-noise ratio at the output is equal to 40 db for PPM modulation, whereas for coded PCM modulation, it is 60 db; this can be seen from Fig. IV.5, in the chapter on restrictions on information.

For a smaller input signal-to-noise ratio, on the other hand – 10 db for example – we obtain

$$[(S/N)_{\text{outpot}}]_{\text{PDM}} = 30\text{ db}$$

and

$$[(S/N)_{\text{outpot}}]_{\text{PCM}} = 10\text{ db.}$$

In this case position modulation seems preferable to coded modulation. In fact, with a choice of bandwidth of $1{\cdot}5f_r$ for the PCM modulation, or

1·275 Mc/s, the bandwidth conditions would no longer be equivalent. In Fig. VI.18 are a series of curves which express the HF bandwidth as a function of the signal-to-noise ratio for various types of modulation. The comparison favours PCM–AM and PCM–FM systems, which occupy a narrower band for a given signal-to-noise ratio.

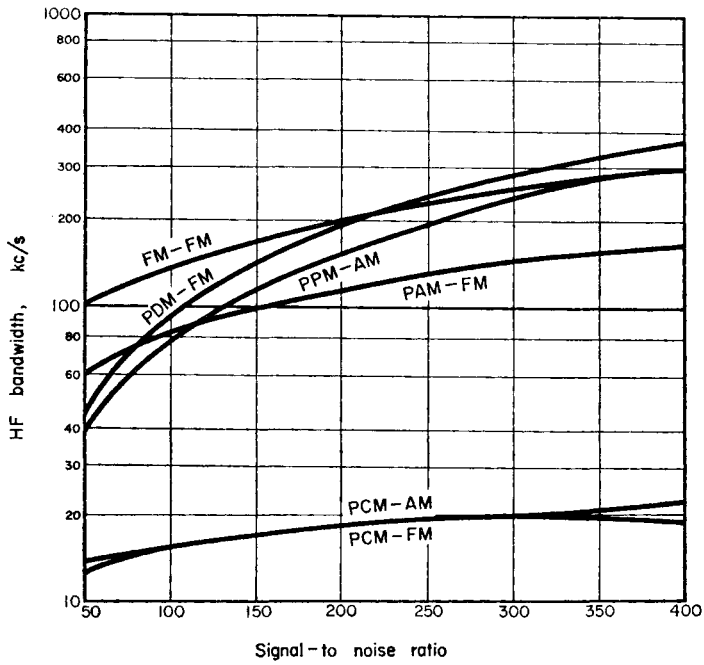


FIG. VI.18. Bandwidth–signal-to-noise ratio characteristic for various types of modulation.[84]

We now consider the problems introduced by jamming. The essential precautions against this disturbance are based upon four points:

– increasing the power to the maximum
– reducing the bandwidth to the minimum
– using the diversity procedure
– coding the information.

Since the information must necessarily be coded, one is tempted to choose the PCM procedure. It is, however, possible to code the channels by agreeing that it shall be the respective positions of several pulses in a well determined sequence which determines the code, and thus to employ the PCM principle less directly.

PCM modulation is undoubtedly the most appropriate procedure for limiting the bandwidth. The anti-jamming problem is an important one particu-

larly where the remote-control of missiles is concerned. Since remote-control commands require only a small amount of information, the information pulses can be time-spread in a sequential form.

Suppose that a response time of 0·3 seconds is sufficient for a remote-control order to be transmitted; with a bandwidth of 200 c/s eight remote-control channels can be employed. In fact, with five digits at our disposal there are 2^5, or 32, possibilities. But a minimum of $2^3 = 8$ digits are necessary to retransmit information. There are thus some digits over, which can be used for coding the information.

Further, if the synchronization pulse is regarded as a supplementary digit, the period of a digit will be $0{\cdot}3/6 = 0{\cdot}05$ seconds which corresponds to a frequency of 20 periods. For the circuits to be compatible with this information, however, a bandwidth of 200 c/s would be selected. The PCM procedure is thus valuable for the retransmission of information with a very small bandwidth.

Where telemetry is concerned, the retransmission must be faithful in terms of the linearity between the measured and detected variation; for this the FM–FM and FM–AM procedures are found to be very satisfactory. The most frequently employed telemetry links at present are FM–AM and FM–FM.

CHAPTER VII

THE GUIDANCE OF MISSILES BY RADIO REMOTE-CONTROL

THERE is a distinction between self-guidance and remote-control guidance of missiles. Remote-control guidance is effected from a control-post which is either stationary or mobile, according to whether the operator directs the missile from the ground or from an aircraft. Self-guidance does not require the intervention of any human agent at all; once the missile has been launched, it steers itself automatically towards its target by means of a self-steering mechanism known as a homing-head. This is a device which detects any error between the missile-target direction and the axis of the missile. This error is applied to a control system which automatically corrects the error in the course.

VII.1. THE POSITION OF REMOTE-CONTROL IN THE NAVIGATIONAL SYSTEM

In order to remotely control a missile towards a well-determined goal, it is essential that its position coordinates, those of the target, and their rates of change, should be known at every instant. Once these parameters have

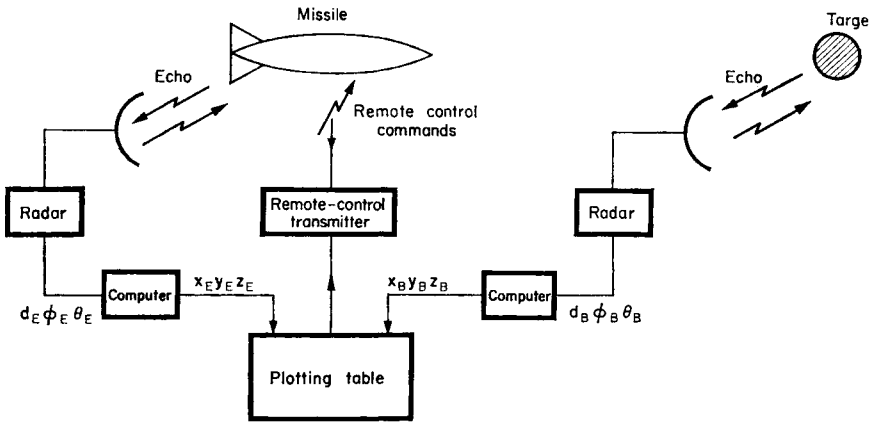


FIG. VII.1. Missile guidance installation.

been ascertained and recorded, an interception can be accomplished by sending remote-control commands which direct the missile towards its goal by actuating its flight controls; the commands may also operate other systems in the missile.

The position and speed of the missile and the target can be plotted either by means of radar, or with an IFF (Interrogation Friend of Foe) device.

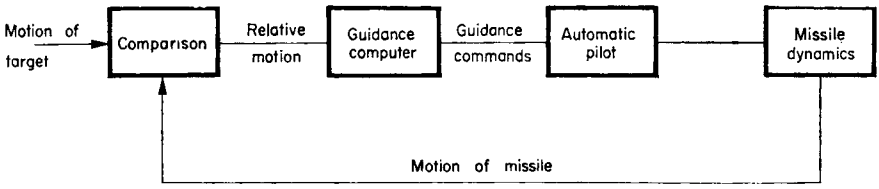


FIG. VII.2. Method of guidance.

The data provided by radar or the IFF can be fed into a computer which transforms the polar coordinates into cartesian coordinates so that the movements of the missile and its target can be displayed on an XY recorder. This display can also be performed electronically in the form of points on the screen of a long-persistence oscilloscope. The altitudes of missile and target can equally well be recorded. A typical complete control installation is illustrated in Fig. VII.2. Once the relative motions of the missile and its goal have been measured, the correction required to ensure an interruption is known. The remote control effects this correction by actuating the missile control systems via the automatic pilot.

The function of the automatic pilot is to transform the remote-control instructions into movements of the missile.

The navigational technique is related to the kinematics of the missile; the way in which the remote control instructions are carried out will depend upon this. We shall distinguish between velocity navigation, which attempts constantly to direct the velocity vector along the line joining the missile to its target and positional navigation, where the coordinates of the missile alone are considered. This second navigational technique is only applicable when the missile is not very fast-moving, and when no great accuracy is required.

By means of remote-control, the behaviour of the missile is decided and imposed upon it from the outside (this is not the case with self-guiding missiles). A more detailed diagram of a guidance system with remote-control is provided in Fig. VII.3.

The calculation of the correction is often performed by a human operator who gives the remote control commands as he judges appropriate to correct the error in the course of the missile. It is often difficult to establish adequate equations for the operator and thus difficulties arise in stability analyses.

The occurrence of oscillations in the complete guidance sequence may in fact depend upon the operator.

The nature of a manual control must, therefore, be compatible with the response-time of the operator. It has been found that a human operator can easily control, for example, the flight behaviour, the speed in normal flying conditions and the landing or retrieval of a target aeroplane, but finds it difficult to follow rapid manoeuvres. When the flight takes place at a constant altitude, for example, it is preferable to provide the automatic pilot with a barostatic probe, or a radio altimeter probe, and to exclude a human intermediary.

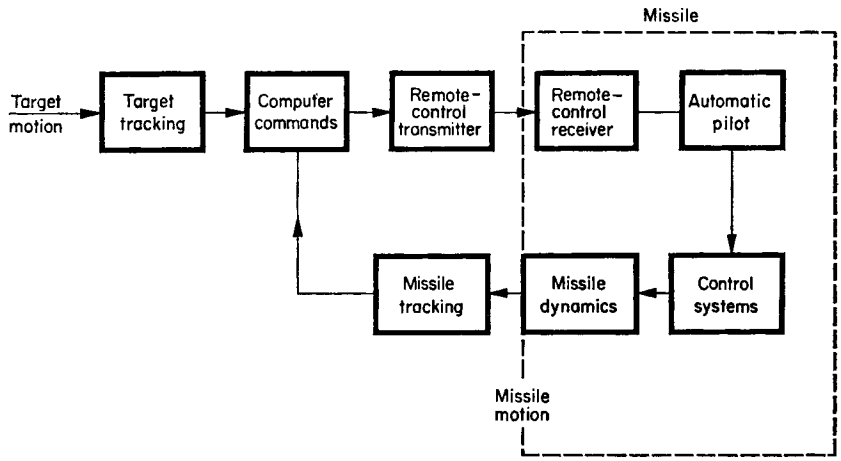


FIG. VII.3. Guidance by remote-control.

So far as the choice of remote-control is concerned, the problems which arise can only be solved subject to considerations of information capacity, jamming, range, noise levels and cross-modulation between the various remote-control channels.

When there is noise, and a velocity—and hence a derivative—is involved, a false guidance order may occur on account of this noise. This perturbation can be suppressed by filtering, but this is an operation which introduces error into the detection of the missile–target direction error.

VII.2. PILOTING BY REMOTE-CONTROL

Since the role of the automatic pilot is important in the navigation procedure, its structure and the ways of guiding a missile will be described before considering the various remote-control commands.

PLATE I. Launching a CT 20.

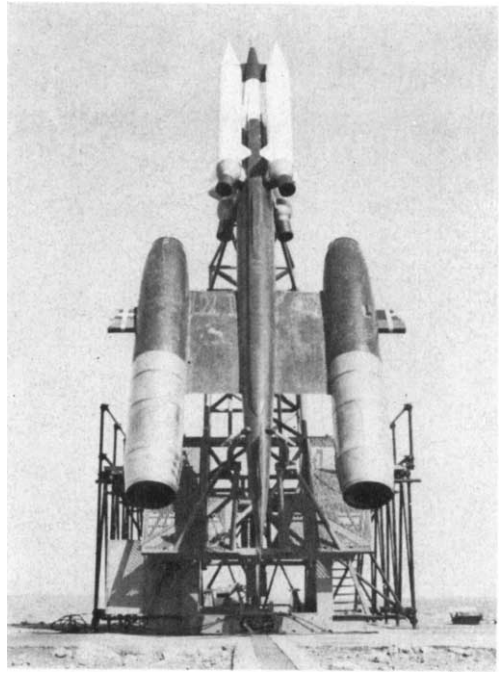


PLATE II. A CT 41 missile on its ramp.

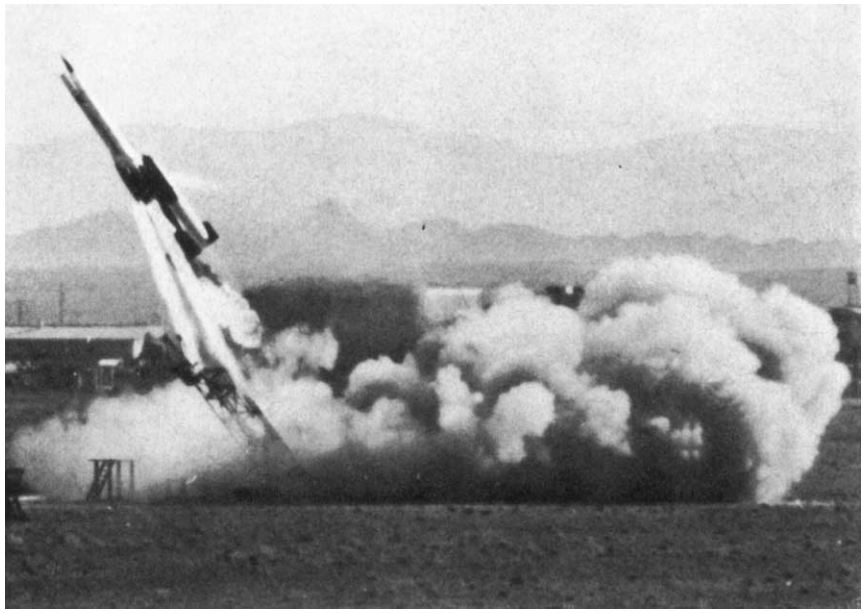


PLATE III. Departure of a CT 41.

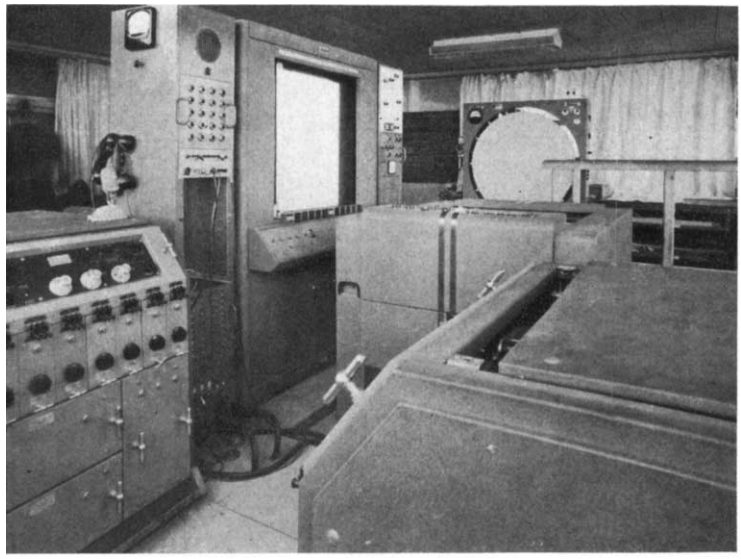


PLATE IV. A fire-control post.

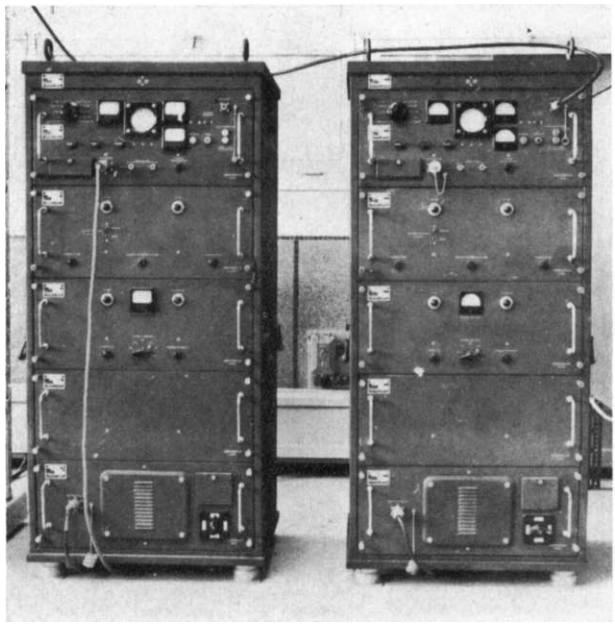


PLATE V. An AM remote-control transmitter.

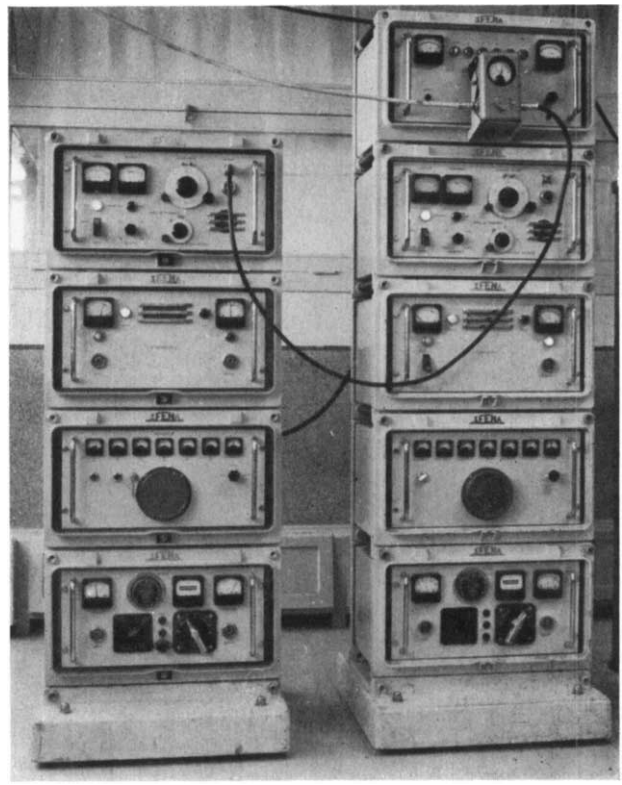


PLATE VI. An FM–FM remote-control transmitter.

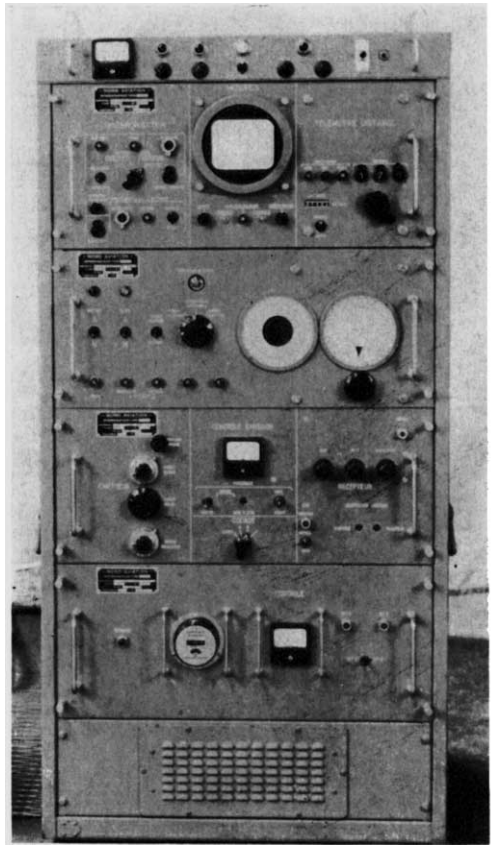


PLATE VII. A LIRNA position-finder.

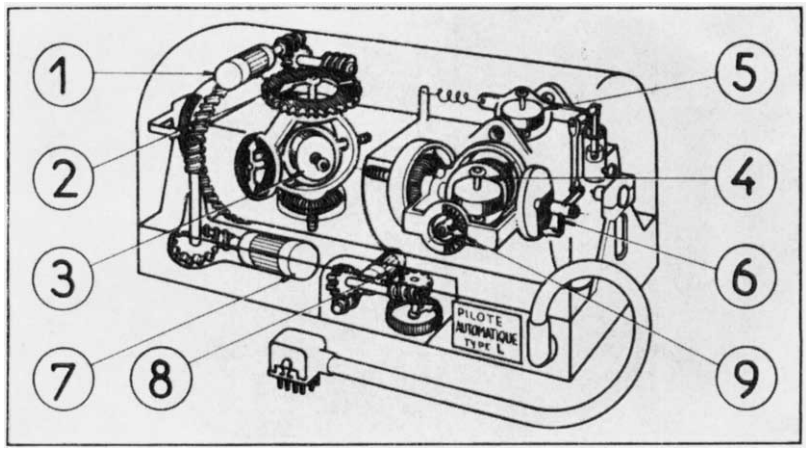


PLATE VIII. The automatic pilot of a CT 20.

(1) motor which is actuated by the "left" or "right" commands of the remote-control and which sets the reference angle. (2) yaw, course disc and potentiometer detector. (3) directional gyroscope. (4) vertical gyroscope. (5) rate gyroscope, indicating the angular velocity of roll $d\varphi/dt$. (6) roll angle detector. (7) cradle control motor. (8) height control setter (reduction geared motor). (9) gimbal potentiometer.

VII.2.1. *The Automatic Pilot*

The remote-control commands are applied to the automatic pilot which transforms the electrical magnitudes into mechanical quantities – alterations to the settings of the flight controls. The latter in turn act upon the flight-behaviour of the missile, and hence upon its trajectory (see Fig. VII.4).

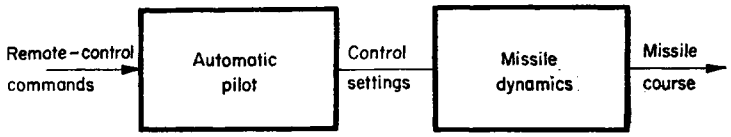


FIG. VII.4. The function of the automatic pilot.

Stabilization must also be introduced into the system, often in feedback form, since otherwise the missile would oscillate about the desired direction.

The angles which define the orientation, and their derivatives, for example, can be determined by means of a gyroscope and a rate gyroscope; and with the aid of a resetting feedback, the correction is applied to the guidance controls.

Further, the measured angle through which the flight controls are moved is compared with the desired angle, in order to ensure that the piloting is

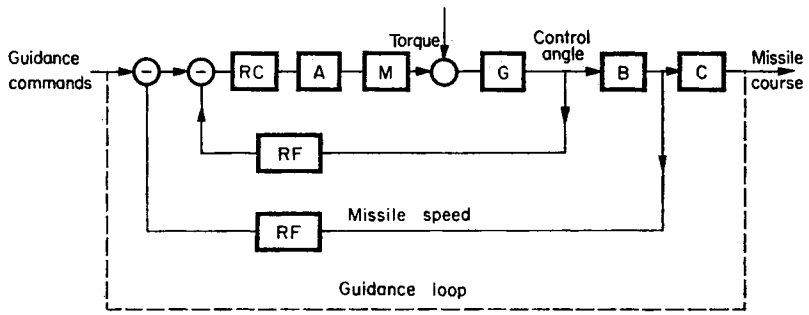


FIG. VII.5 Diagram of the automatic pilot.

RC: stabilizing network; A: amplifier; M: motor; G: flight controls; B: movement of the missile about its centre of gravity; C: movement of the centre of gravity of the missile; RF: transducers.

independent of random perturbations which might occur. We must clearly distinguish between the control loop for positioning the controls, the stabilization loop and the complete guidance loop which includes the operator. The overall arrangement of an automatic pilot is shown in Fig. VII.5.

Three signals have, therefore, to be fed into the automatic pilot:
– the guidance instructions themselves
– the feedback from the flight controls
– the missile reference and stabilization signals.

It should be noted that when a missile banks durin flight, it loses altitude. An automatic "climb" device is therefore incorporated in the "height" loop which behaves like the function $A(\sec\varphi - 1)$, in which φ is the inclination of the missile.

The missile can, in general, be navigated in one of two ways – position and velocity. Positional navigation requires that the turning angles – for height and inclination – are proportional to the remote-control signals

$$\beta = K_1 S_p,$$

$$\varphi = K_2 S_I,$$

in which S_p and S_I represent the remote-control signals which correspond to height and inclination.

For velocity navigation

$$\frac{\mathrm{d}\beta}{\mathrm{d}t} = K_1 S_p,$$

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = K_2' S_i,$$

This introduces an integration into the guidance loop, and as a result a phase-change of $\pi/2$, which must be compensated for.

VII.2.2. *The Remote-control Commands and their Channel Allocation*

The coordinates of the missile (M) and its target (T) are known at every instant with respect to a fixed station, O. The problem, therefore, is to remove any difference in direction between the vectors OM and OT, and to ensure that the function $|MT|$ is always a decreasing function of time. The problem is to decide what remote-control command is to be given, and for how long, in order that the guidance law be followed perfectly.

There is a relation connecting the angular errors ε_y and ε_z with respect to the reference direction in the aircraft (differences in direction and position, or skew, between the axis of the missile and the missile-target direction) and the angular errors in direction and position between the vectors OM and OT and axes fixed with respect to O. It is in general

$$\varepsilon_y = f(\Delta\theta, \Delta\alpha)$$

$$\varepsilon_z = g(\Delta\theta, \Delta\alpha).$$

If the functions f and g are known, the problem is simple, and in general the times for which the elementary commands used in remote-control (dive, climb, left, right) are applied are proportional to ε_y and ε_z.

Calculation of the functions f and g is a relatively complicated procedure, as not only are the angular errors $\Delta\theta$ and $\Delta\alpha$ involved, but also the distances $|OM|$ and $|OT|$ and the control velocities v_M and v_T of missile and target. It is the task of the control computer to solve these equations. The fact that the problem is complicated automatically implies cumbersome equipment. This computer must therefore be situated on the ground.

When the missile is of the Air–Air type, the navigation is performed from the launching aircraft, and the volume problem is formidable. An interesting solution consists of re-transmitting the position and direction errors of the missile back to the ground by telemetry; these errors are measured with respect to the vertical and horizontal directions, defined by a gyroscopic reference. The errors themselves can be measured by means of gyroscopes. The values of the correction which must be applied by remote-control are thus directly obtained. If the remote-control and the telemetry transmission are linked directly by a control system which gives the commands, the resulting system is known as "automatic remote-control".

The telemetry equipment may also consist of a television unit, with the camera placed in the nose of the missile. A UHF transmitter transmits the image which is formed, and the operator thus steers the missile by direct vision; this is useful for missiles of limited range.

The remote-control commands which are required are generally of two kinds; guidance commands in the proper sense, such as dive, climb, right and left, together with supplementary commands, to increase or reduce speed, to recover from some manoeuvre, to record or to take photographs, to detonate explosive charges, to make smoke, or to put some special device into action.

Some remote-control commands are mutually exclusive, dive and climb for example. Further, all the supplementary remote-control orders will not be transmitted simultaneously. Due to this fact, a large number of remote-control instructions can be made available in the form of combinations of independent primary commands. Suppose there are six remote-control channels available (as is the case of the remote-control of the Nord Aviation CT 41). Two independent groups of orders are formed, A and B. Group A covers the channels 1, 2, 3 and 4, while B covers 5 and 6. A command from group A and another from group B can be transmitted simultaneously. The allocation of the remote-control commands can be described in the following way: the symbolic product of the commands corresponds to the logical combination "AND", whereas their sum corresponds to "OR".

The command to "keep going" is thus produced at the transmitter in the following way:

(i) in the absence of any other commands it can be produced automatically by applying the combination (3).(4) + (5).(6)

(ii) when a command from group A or group B is present, maintain course is automatically obtained by applying (5).(6) or (3).(4) respectively.

At the receiver, the "keep going" message is characterized by the presence of a command from any channel, and hence by the combination (1) + (2) + (3) + (4) + (5) + (6).

Group A		Group B	
(1)	supplementary	(5)	right
(2)	measurement	(6)	left
(3)	camera	(5).(6)	keep going
(4)	flame-thrower		
(3).(4)	keep going		
(2).(4)	increase speed		
(2).(3)	reduce speed		
(1).(4)	dive		
(1).(3)	climb		
(1).(2)	recover		

More generally, suppose there are N remote-control channels. If n is the number of channels affected by a command, then the maximum number of remote-control commands which can be attained is $M = C_N^n$ (the number of ways of selecting n channels from among N possible channels). M is a maximum for $n = N/2$ when N is even and for $n = (N + 1)/2$ when N is odd.

If $N = 6$ and $n = 3$, the maximum number of remote-control commands which could be achieved would be $M = 20$.

VII.3. DECODING THE REMOTE-CONTROL COMMANDS

The information is transmitted by radio and always appears at the output of the remote-control receiver in the form of an electric signal. These voltages therefore have to be transformed into a quantity suited to its subsequent use; this may be electrical (the circuit which detonates the explosive charges) or mechanical.

The problem is only outlined here, since the object of this book is primarily to consider the radio part of the link. It is interesting however to know something of the direct applications.

VII.3.1. *The Servomotor that actuates the Height Control*

Figure VII.6 shows such a device. The mean current through the shunt motor is determined by the difference between the voltage associated with

the position of the control and the electrical signal at the output of the receiver. The tacho-generator provides damping.

Alternatively the system shown in Fig. VII.7 can be used; it is quite standard, and was employed in the German HS 293 missile.

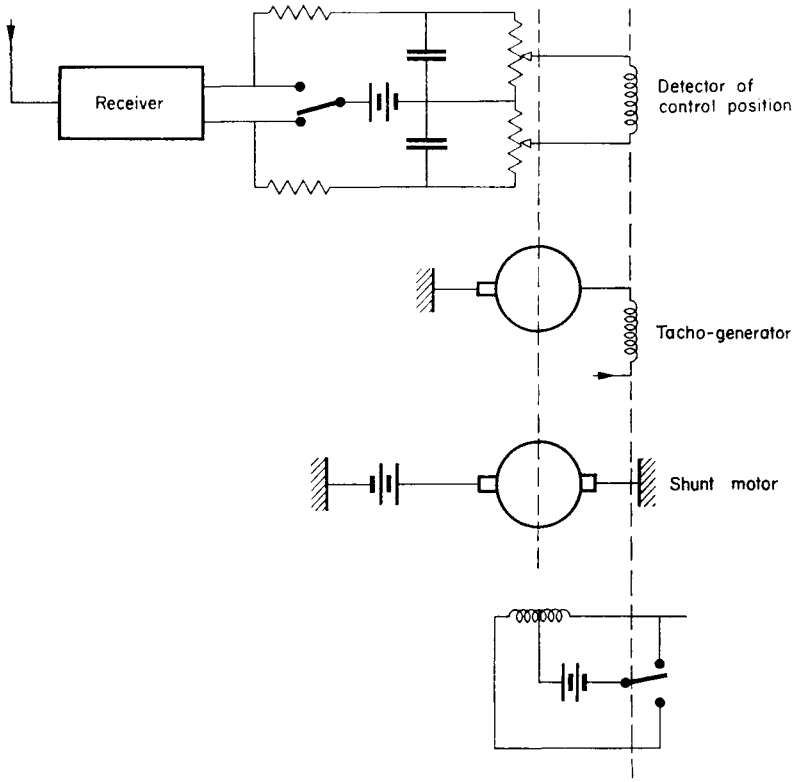


FIG. VII.6. System for setting height controls.

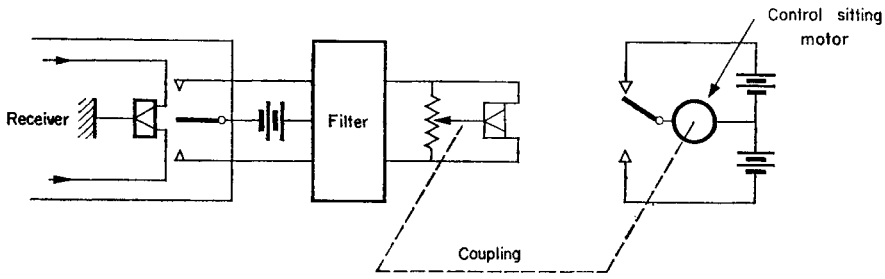


FIG. VII.7. Alternation system.

VII.3.2. *Description of the Control Equipment in the Nord-Aviation Target Missile, CT 20*

The automatic pilot consists of two gyroscopic stabilizers: one against yawing, the other against rolling and pitching. The whole device consists of a central gyroscope unit, shown in Plate VIII, which provides three reference directions; these correspond to the axes about which the craft yaws, pitches and rolls.

Stabilization against Rolling (see Fig. VII.8 and Plate VIII). A contact, which is linked to the roll rate gyro and is sensitive to roll, is constrained to move across a disc attached to the vertical gyro. The disc has two sectors, one insulated and one conducting.

The contact-disc circuit is thus affected by the appearance of roll and this is used to modulate the mark–space ratio of the flapping aileron control surfaces on the missile wings, so as to correct the roll.

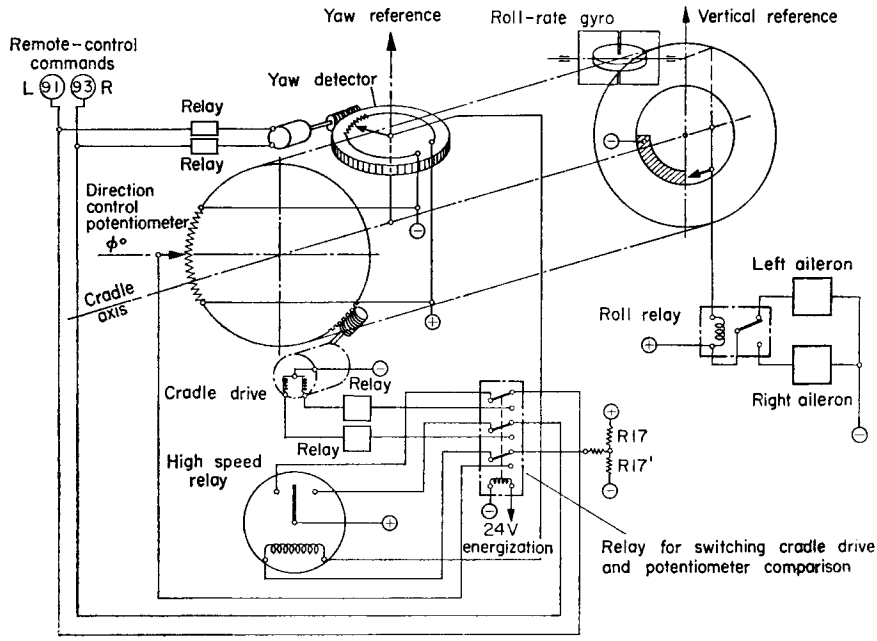


FIG. VII.8. Diagram showing course control, roll and yaw stabilization, by the automatic pilot in the Nord-Aviation CT 20.

Stabilization against Yawing (see Fig. VII.8 and Plate VIII). During straight level flight the wipers of the yaw detector and the directional control potentiometers are at their midpoints; as a result the high speed relay is

unenergized. If a perturbation causes the missile to yaw, the detector is offset and the relay is energized, in the appropriate sense. This drives the cradle motor until the control potentiometer produces a voltage equal and opposite to that of the detector. This moves the roll detector disc which actuates the flapping ailerons in a manner which corrects the yawing.

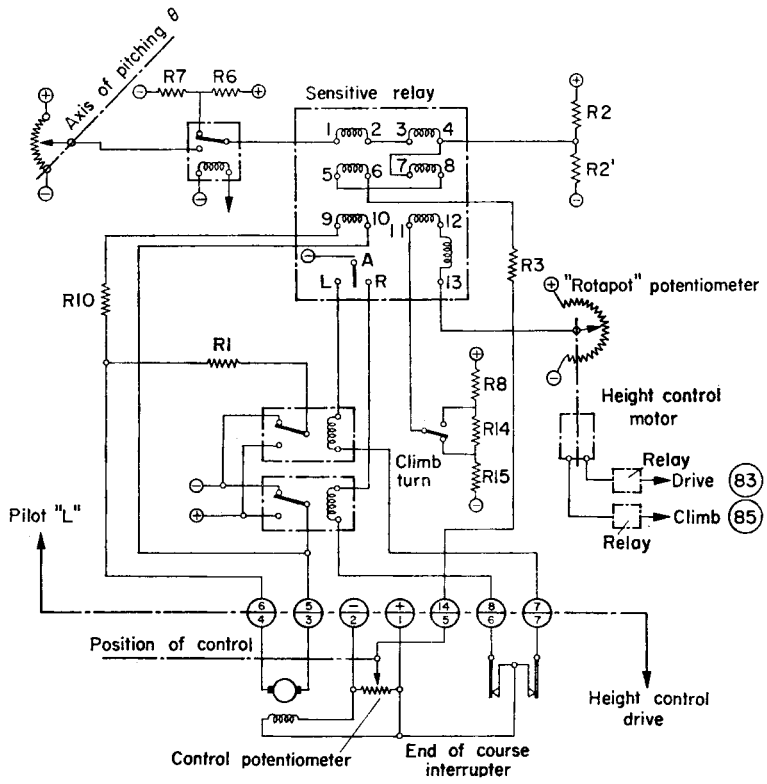


FIG. VII.9. Diagram showing height control and pitch stabilization, by the automatic pilot in the Nord Aviation CZ 20.

To carry out a change of course by remote-control, the command operates one of the power relays associated with the motor which rotates the yaw detector potentiometer card. This sets a new yaw reference angle proportional to the duration of the command; the missile takes up its new course at this reference, with roll stabilization.

Stabilization against Pitching (see Fig. VII.9). Equilibrium corresponds to the central position of the wiper of the height control potentiometer. In this position, no current flows through the windings 11–12 and 12–13 of the sensitive relay.

The fields produced by the windings 1–2 and 3–4, which are supplied by the potentiometer which detects pitching, are equal and opposite in sense to those in the windings 5–6 and 7–8; the latter are supplied by the control potentiometer which is connected to the position of the controls. A magnetic equilibrium of the relay is thus established. An external perturbation is all that is needed to destroy this equilibrium and this results in an appropriate adjustment of the height controls. This unbalance can equally well be produced by applying the order dive or climb, as either of these orders sets the height-control motor in action, and alters the fields in the windings 11–12–13. The missile then settles down about the new reference direction which the position of the slider of the height potentiometer defines.

VII.3.3. *Control by Mark-space Variation*

Consider the series of rectangular pulses at the output of the remote-control receiver shown in Fig. VII.10.

The repetition period, θ, remains fixed but the intervals θ_1 and θ_2 vary as a function of the remote-control commands which are applied ($\theta = \theta_1 + \theta_2$).

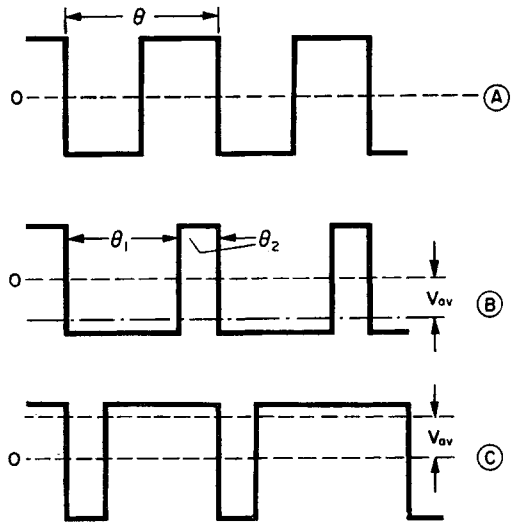


FIG. VII.10. Mark-space variation.

From the diagram, it can be seen that if $\theta_1 = \theta_2$, the mean value of the signal is zero, whereas if $\theta_1 > \theta_2$, it is negative. This variation can be used to actuate elements such as the jet interceptors of a missile, or to turn a motor in one sense or the other according to the respective values of θ_1 and θ_2.

Application to Flapping Aileron Control. At the edges of the two wings of remote-controlled aeroplanes are to be found two mobile flaps (these may take the form of flaps mounted and moving perpendicular to the wing surface and housed in slots, in which case they are termed "spoilers"; alternatively trailing edge flaps, or ailerons, can be used which rotate from "up" to "down"). Their mode of operation is illustrated in Fig. VII.11.

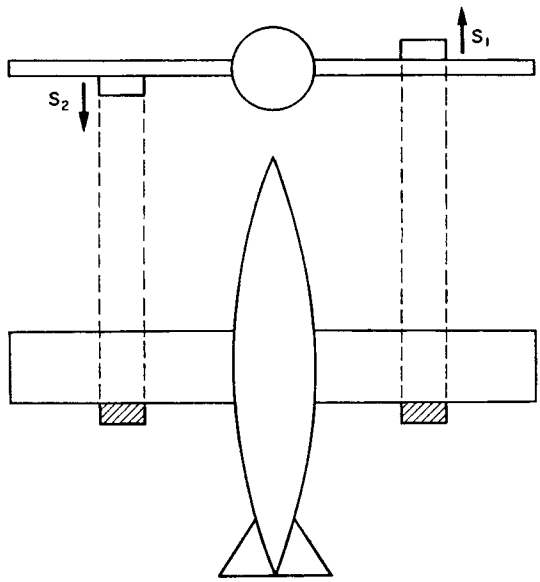


FIG. VII.11. Missile flapping aileron controls.

When S_1 is in the "down" position, S_2 is automatically in the "up" position. The slipstream is deflected in opposite senses at S_1 and S_2, which results in a couple about the yaw axis; the lifts of the two wings are altered, and the aeroplane therefore banks.

The ailerons S_1, S_2 flap regularly during straight line flight, which means that $\theta_1 = \theta_2$; if the periodicity is thrown out of balance, however, if for example, θ_1 is greater than θ_2 – the missile banks (in the case illustrated in the diagram) to the left.

Application to Control of the Jet Deflectors. In the case of missiles spinning about their axis, four jet deflectors in the form of shutters are provided; these act symmetrically at the jet chamber orifice in such a way as to obstruct it to a greater or lesser extent. These shutters alternately open and close and are coupled together in pairs. For the height channel, the "bottom" deflector is closed when the "top" deflector is open, and vice versa. The principle is identical for the direction channel. The alternating motion is performed at a frequency of 10 c/s. The command itself is defined by the ratio $x = \theta_2/\theta_1$

of the periods of the two alternating systems. The quantity x can vary at most from 0 to 1.

As Fig. VII.12 shows, electromagnets are used to actuate the deflectors.

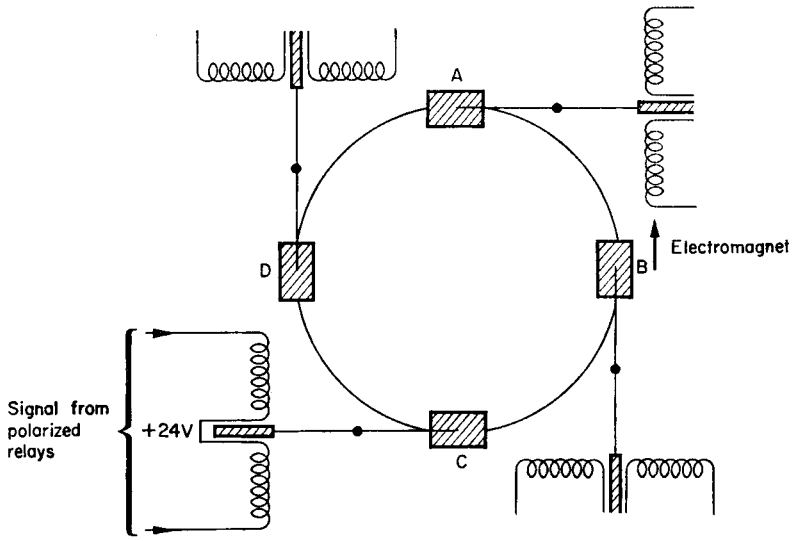


FIG. VII.12. Layout of jet deflector.

When the missile is spinning about its axis, the remote-control commands pass through the intermediary of rings connected to a gyroscope. The latter commutates the deflector to the corresponding channel while the missile is rotating. It carries out a form of change of coordinates related to the rotation of the missile about itself.

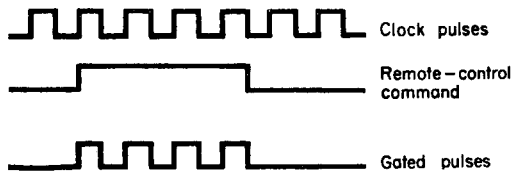


FIG. VII.13. Gating of pulses.

Control by Pulse-counting. This type of control is particularly well-adapted for accurate control over some elements, for example, the rotation of a stepping motor. As a concrete example let us suppose that the course set by an automatic pilot is altered by rotating a reference direction (the wiper of a potentiometer). The maximum required course angle is say 5°, with an accuracy of 0·1°, and a stepping motor with 50 positions is employed. Passage

from one position to the next corresponds to 0·1°. To set an angle of $\alpha°$ by radio, requires the transmission of 10α pulses. The remote-control command is applied by operating a relay which closes while the time for 10α pulses elapses. The command pulse is made to gate a clock pulse train, and the resulting pulses are applied to the stepping motor (see Figs. VII.13 and VII.14).

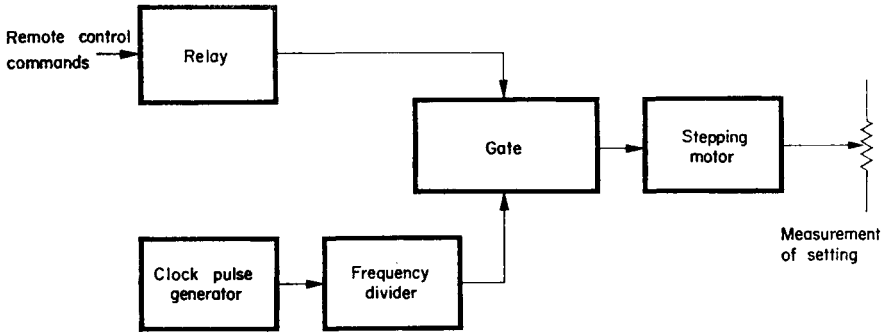


FIG. VII.14. Remote-control of stepping motor.

VII.4. EXAMPLES OF RADIO REMOTE-CONTROL SYSTEMS

VII.4.1. *An Example of FM–FM Remote-control*

VII.4.1.1. *General Remarks*

By way of example, a double frequency modulation remote-control system with six independent channels will be described. The block-diagram of the transmitter is given in Fig. VII.15.

The elementary orders 1, 2, 3, 4, 5 and 6 connect the LF oscillators O_1, O_2 and O_3 into the circuit; these give signals at frequencies F_1, F_2 and F_3 which modulate the subcarrier waves O_{SP_1} and O_{SP_2}. These two modulated subcarrier waves, with frequencies f_{S_1} and f_{S_2}, are mixed, and the resulting signal in turn modulates the carrier wave. After amplification, the output stage delivers the necessary power at the transmitting aerial. Reception at the missile is effected in the following way (Fig. VII.16).

The RF signal is fed to an input amplifier; after mixing with a local oscillator, it is amplified in an IF stage. The composite signal constituted by all the frequency-modulated subcarrier waves is then re-established by means of a discriminator. The subcarrier waves are separated by filtering, and the LF channels appear at the output of a second discriminator. After filtering, amplification and detection, these LF signals control the relays in the order 1, 2, 3, 4, 5, 6.

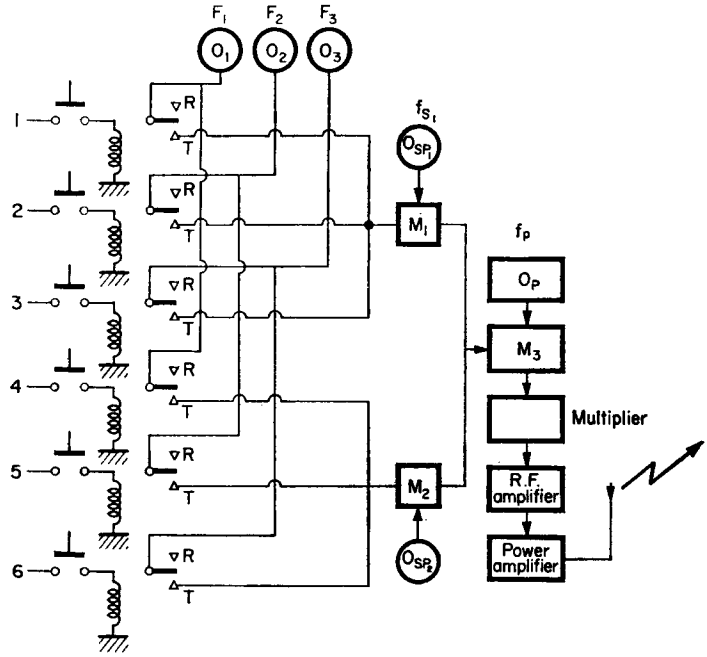


FIG. VII.15. Schematic diagram of an FM–FM remote-control transmitter. O_1, O_2, O_3: LF oscillators; O_{SP_1}, O_{SP_2}: subcarrier wave oscillators; M_1, M_2: subcarrier wave modulators; O_P: carrier wave oscillator; M_3: carrier wave modulator

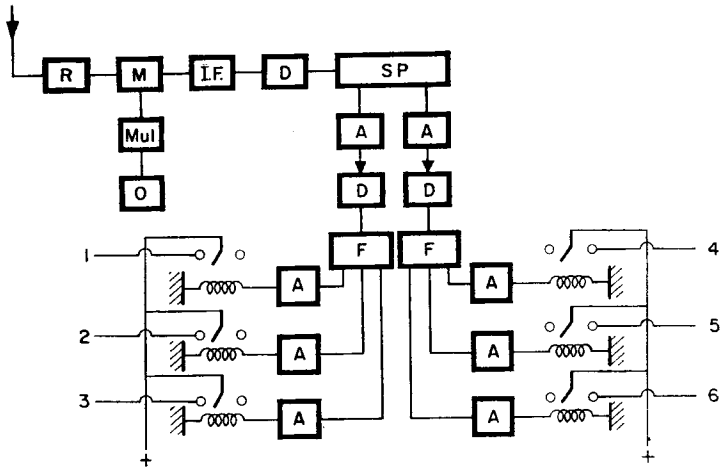


FIG. VII.16. Schematic diagram of a FM–FM remote-control receiver. *RF*: RF stage; *M*: mixer; *Mul*: multiplier; *O*: oscillator; *IF*: IF amplifier; *D*: discriminator; *SP*: subcarrier wave filter; *A*: amplifiers; *F*: LF filter.

To increase the number of remote-control channels, starting from these six elementary orders, all the possible combinations of these orders can be used. An example is given in the table below.

TABLE VII.1.

Keep going	1
Recovery: absence of the order	1
Dive	2
Climb	3
Left	4
Right	5
Camera	2 + 6
Make smoke	3 + 6
Increase speed	4 + 6
Decrease speed	5 + 6

VII.4.1.2. *Devices for Modulating the Subcarrier Waves*

Two LF oscillators are shown diagrammatically in Fig. VII.17: one is transistorized, the other employs valves.

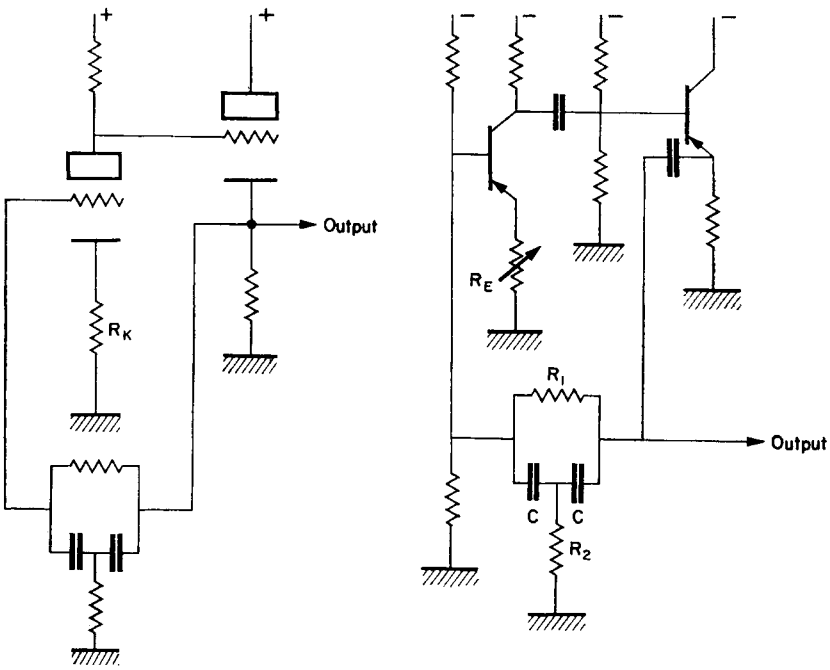


FIG. VII.17. LF valve and transistor oscillators, using a resistance–capacitance bridge.

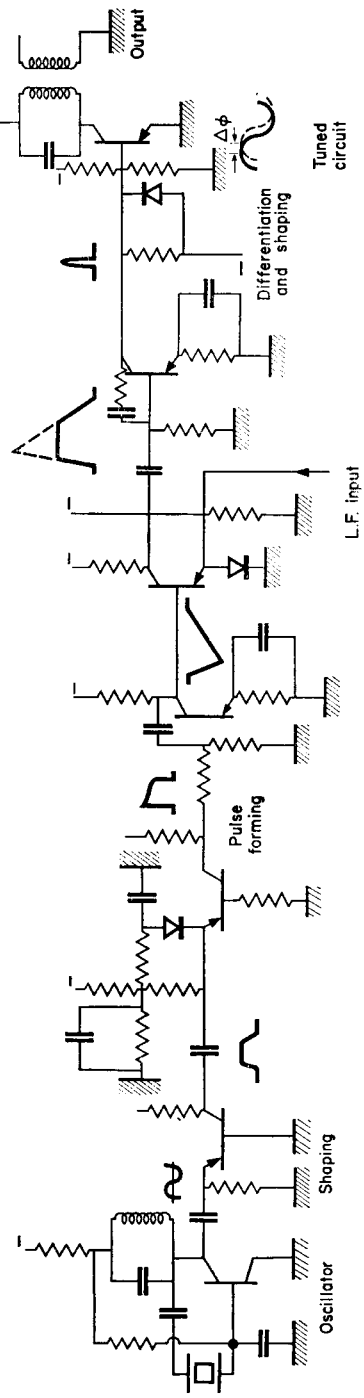


FIG. VII.18. Serasoid modulation.

The frequency is adjusted by altering the capacitor C and the resistors R_1 and R_2. Signals with only low distortion can easily be obtained, and such a device avoids the use of inductors which are usually bulky. The resistor R_3 in the emitter circuit is adjusted to give the desired output voltage.

The phase of the subcarrier is modulated by means of a serasoid device. A quartz clock generator provides a frequency of 100 kc/s which is then shaped to give a control pulse which is converted into a saw-tooth. When there is no LF modulation, this pulse is limited at its midpoint. When an LF signal is applied the saw-tooth is limited at a different height, so that the limiting instant is displaced along the time axis. After differentiation, a phase-modulated pulse is obtained (see Fig. VII.18).

This pulse passes into a filter giving a sinusoidal wave modulated in phase at 100 kc/s. On mixing with a signal at 100 kc/s $+ f_{S1}$, the subcarrier wave with frequency f_{S1} is obtanied. An identical procedure produces the subcarrier wave of frequency f_{S2}. These two waves are mixed, and modulate the carrier. The linearity of serasoid devices is excellent.

VII.4.1.3. *Modulation of the Carrier*

The production of a phase-modulated carrier wave is a standard procedure. It is, however, worth mentioning the present trend in the use of certain semiconductor devices such as "varicaps". These elements are basically silicon junction slides whose capacitance varies appreciably with the applied bias. This is therefore a practical means of modulating the frequency of an oscillator, by varying the tuning capacitance.

The capacitance, C, is expressed as a function of the applied voltage, V, by

$$C = k(V + V_0)^{-n}$$

in which V_0 is the contact potential of the junction of the order of 0·3 to 0·9 volts; k is a constant which is a function of the resistivity of the semiconductor and the contact surfaces. For an abrupt junction, $n = \frac{1}{2}$. A circuit of a frequency modulator using a varicap is given in Fig. VII.19.

In Fig. VII.20 is plotted the variation of the capacitance C and of the quality factor $Q = 1/\omega RC$ as a function of the bias voltage potential.

From Fig. VII.21, it can be seen that the quality factor passes through a maximum when the modulation frequency is 100 kc/s.

A modulation device using a reactance transistor is shown in Fig. VII.22.

The transistor $T1$ oscillates at the frequency of the quartz oscillator Q, and the modulation is applied to the transistor $T2$. The LF signal alters the bias of $T2$, and this modifies the capacitance appearing across the LC circuit in the collector of $T2$.

The signal at the output of the device passes into a series of frequency multipliers. Once the carrier frequency has been established, a "driver" stage controls the power amplifier which supplies to the transmitting aerial the power nevessary to provide the link.

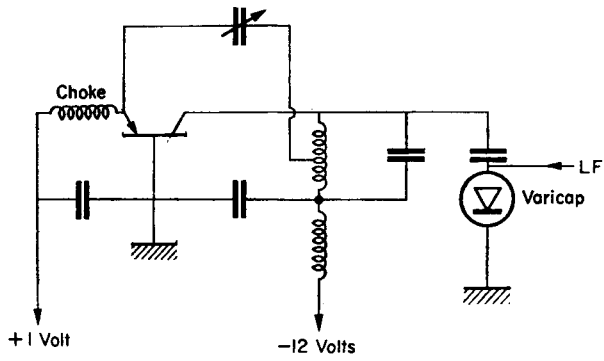


FIG. VII.19. FM modulation, using a varicap.

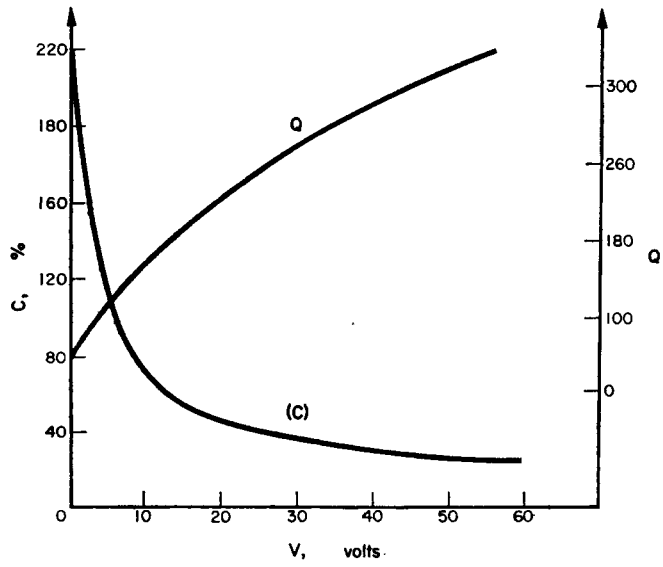


FIG. VII.20. Variation of Q and capacitance of varicap as a function of the bias.

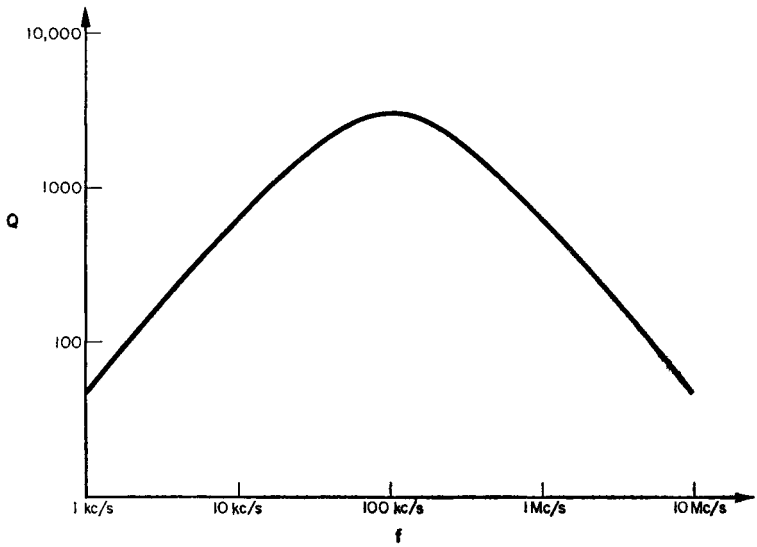


FIG. VII.21. Variation of Q of a varicap as a function of the frequency.

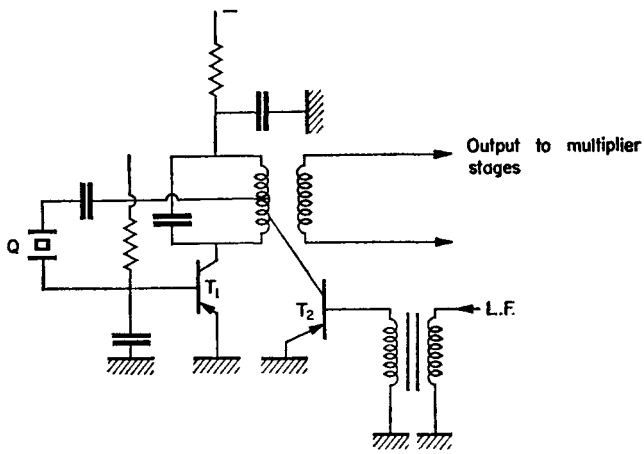


FIG. VII.22. Transistorized reactance modulator.

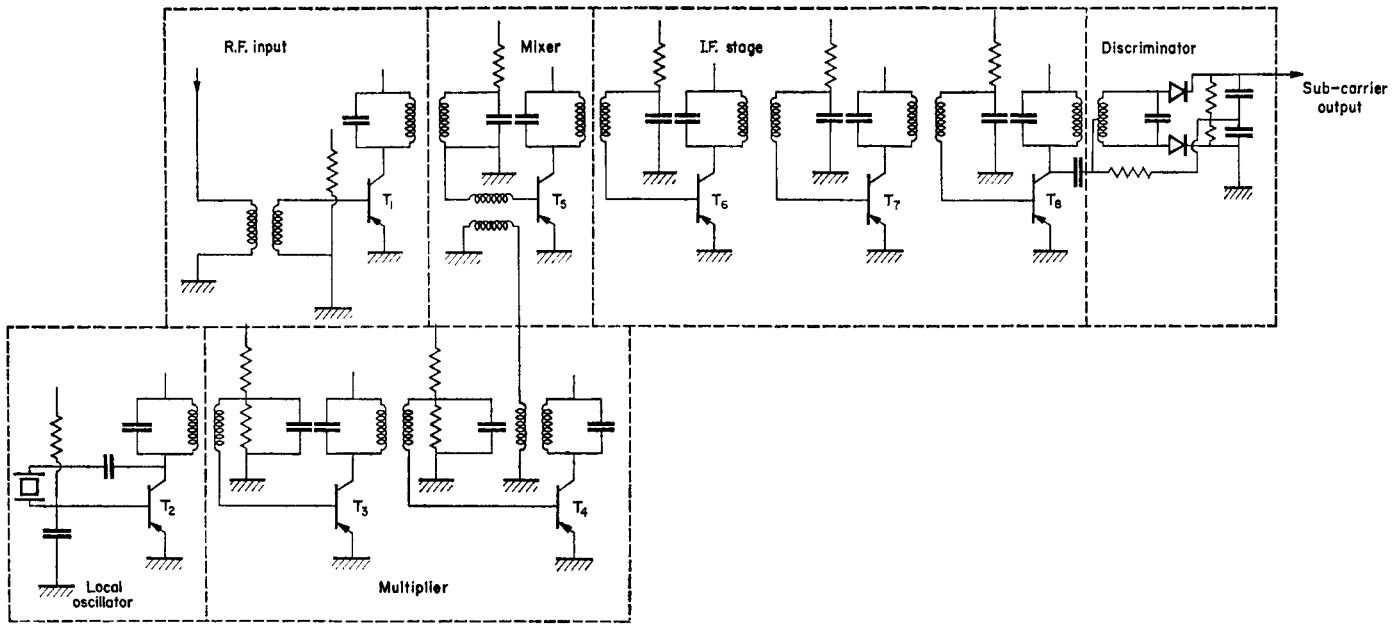


FIG. VII.23. The RF–IF stages of a transistorized FM–FM remote-control receiver.

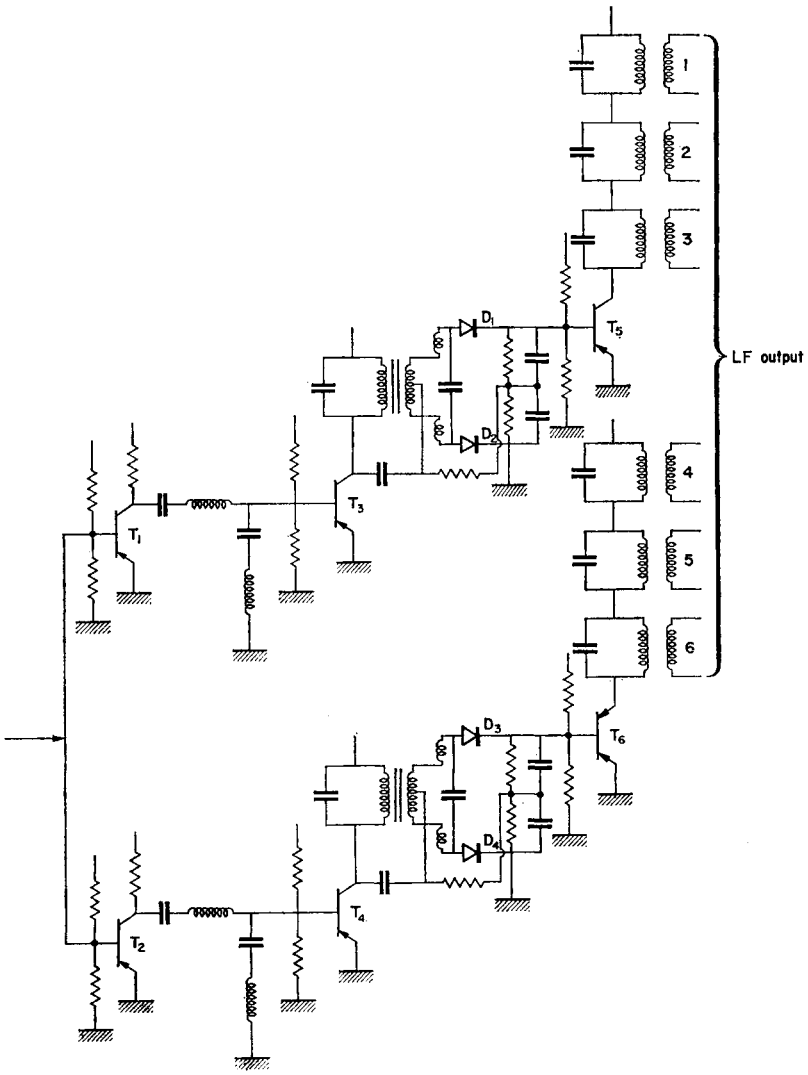


FIG. VII.24. The circuit for demodulating the subcarrier waves.

VII.4.1.4. *Receiving equipment*

The incoming signal is picked up by the receiving aerial. The input stage consists of a transistor $T1$ and its accompanying circuitry (see Fig. VII.23). The local oscillator $T2$ provides a signal at the collector of the transistor which is applied to the two multipliers $T3$ and $T4$. This signal is mixed with the RF input signal at the base of the transistor $T5$ whose collector circuit is tuned to the IF frequency. The IF amplifier consists of the stages $T6$, $T7$ and $T8$. The base-emitter bias of $T8$ is set to limit the signal before applying it to the Foster–Seeley discriminator shown in the figure.

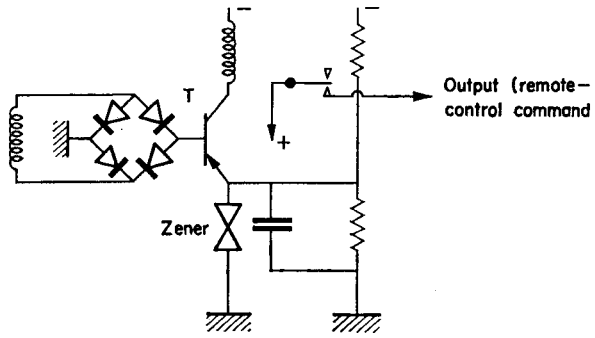


FIG. VII.25. LF detection, production of a remote-control command.

A complex signal consisting of the two subcarrier waves f_{S1} and f_{S2} appears at the output of the RF–IF units. In Fig. VII.24, the method of separating the subcarrier waves and detecting the LF signal is shown.

The input signal is applied to the transistors $T1$ and $T2$ simultaneously. At the output of these two transistors the subcarrier wave is filtered by LC circuits so as to reduce to a minimum the subcarrier frequency of one channel in the other.

The phase discriminator circuits are in the collector circuits of the transistors $T3$ and $T4$. Their outputs consist of complex signals composed of a mixture of the LF frequencies. The signals are applied along each channel to the amplifiers $T5$ and $T6$ respectively. In the collector circuit of these two transistors are three circuits tuned to the three LF frequencies F_1, F_2 and F_3 respectively.

In Fig. VII.25, the detection of LF frequencies is shown schematically together with the method of producing remote-control commands. The LF signal is rectified in a diode bridge–circuit, and the steady voltage which results is used to lift out or cut off the transistor T which is normally biased back by the Zener diode in its emitter circuit. As soon as the steady voltage exceeds

the triggering threshold defined by the Zener diode, the transistor T passes a current through the winding of the relay, and a remote-control command appears at the output.

VII.4.2. *An Example of Digital Remote-control*

VII.4.2.1. *General Remarks*

The remote-control instructions are in the form of a succession of binary words, which are transmitted cyclically. Each order is therefore characterized by a digital word, which can be decomposed into two syllables: address and command. The "address" syllable indicates the component at which the order itself is directed. With a five-digit address, for example, thirty-two different combinations are available. To each of these address digits can be attached additional digits as a check upon the address which is transmitted. When the instructions are of the "yes–no" type, the command syllable consists of two digits, either 11 or 00, which indicate that the command is or is not to be executed. A progressive type of remote control can be obtained by including enough digits in the "command" syllable, making due allowance for the accuracy of the order.

VII.4.2.2. *Transmission*

Since the response-time of an order is inversely proportional to the frequency of the digital rhythm, it is advantageous to speed up this rhythm.

The whole cycle consists of the sections shown below:

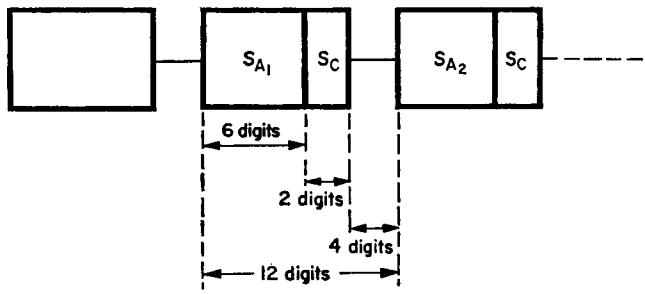


FIG. VII.26.

Since the duration of the interval is fixed at four digits, for a digital rhythm of 100 kc/s, the transmission time of an order is 120 μs (12 × 10 μs).

The signal which is obtained is of the PCM type, and can be used to modulate either a carrier or a sub-carrier in PSK (Phase Shift Keying) or in

FSK (Frequency Shift Keying). The transmitter is shown diagrammatically in Fig.VII.27.

Since the various orders are transmitted cyclically, the corresponding addresses can conveniently be commutated. This can be done with the aid of a ring counter which governs the address store. The digits which form the address are provided in parallel form. Each address, and hence each stage of the ring counter, corresponds to a particular command.

An order is conveyed to the transmitter by closing a contact on the corresponding input line. This creates a potential across this line. When this potential coincides with the crest originating in the corresponding stage of the counter, the command digit is triggered off. To obtain the PCM signal, we have only to perform a series-parallel transformation of the digits. The PCM–

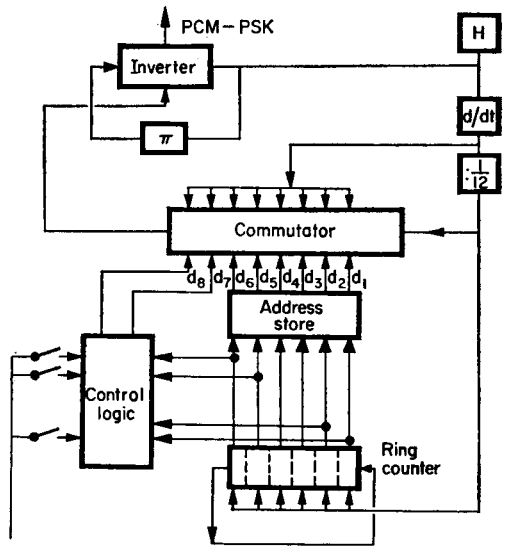


FIG. VII.27. The general layout of a digital remote-control transmitter.

PSK signal which is obtained is applied to the HF transmitter which eventually provides a PCM–PSK–FM signal on a 200 Mc/s carrier.

VII.4.2.3. *The Receiver Unit*

This consists of VHF superheterodyne receiver together with a discriminator, which gives the PCM–PSK signal at its output. The PCM signal in the sub-carrier is detected by comparing the phase of the PCM–PSK signal with that of a clock with the same phase and frequency as the digital rhythm.

The data are detected by correlating the remote-control codes, which have been transmitted, with the codes produced locally; the series PCM signal is converted into parallel form beforehand, as indicated in the figure.

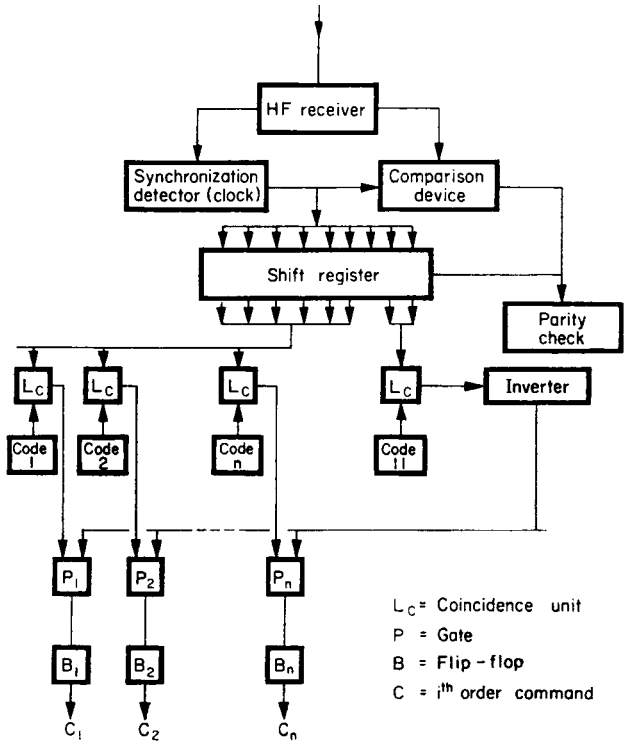


FIG. VII.28. The layout of a digital remote-control receiver.

VII.5. OPTIMUM FILTERING FOR MISSILE GUIDANCE

The guidance of a missile requires the accurate measurement of the parameters "position" and "speed". In addition, delays caused by filtering the data must be minimized. Because these two requirements are contradictory—accuracy often entails considerable filtering—it is necessary to find an acceptable compromise between them.

Further the missile must be suited to the target; its behaviour will therefore depend upon that of its goal. This is characterized by its velocity, altitude, and the properties associated with its motion (e.g. lateral acceleration).

The useful signal provided by the tracking radar is subject to random fluctuations (echo fluctuations, produced by changes in the equivalent surface area of the target, and the phenomenon of scintillation or glint of the phase

centre). It is thus necessary to consider the frequency spectra of the signal and the noise. If the signal spectrum does not coincide with that of the noise, the signal can easily be distinguished from the noise by simple bandpass filtering.

If, on the other hand, the two spectra coincide, there is no way of filtering off the signal unless its shape is known beforehand.

In practice, we say that the filtering of the signal is optimized when the signal-to-noise ratio is a maximum.

Wiener's method can be employed to attain optimum filtering, assuming that the guidance equations of the missile can be considered linear.

VII.5.1. *Optimization of a Linear System*

Consider the linear system of Fig. VII.29. The functions $e(t)$ and $s(t)$ are the input and output signals; their Laplace transforms are $E(p)$ and $S(p)$ respectively, and their frequency spectra are $\varphi_E(\omega)$ and $\varphi_S(\omega)$. The spectra φ_E and φ_S, and $F(p)$, are related by

$$\varphi_S(\omega) = |F(j\omega)|^2 \varphi_E(\omega).$$

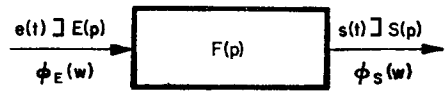


FIG. VII.29. Linear system.

Suppose that noise $n(t)$ is superimposed on the input signal. If $N(p)$ is the Laplace transform of $n(t)$, we have

$$F_1(p) = -\frac{S(p)}{E(p)}$$

$$F_2(p) = \frac{S(p)}{NC\omega}$$

or

$$\varphi_1(\omega) = |F_1(j\omega)|^2 \varphi_E(\omega)$$

and

$$\varphi_2(\omega) = |F_2(j\omega)|^2 \varphi_N(\omega).$$

The parasitic term is therefore contained within the spectrum $\varphi_2(\omega)$. Its mean square value is

$$\overline{\varepsilon^2} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} \varepsilon^2(t)\, dt,$$

or

$$\overline{\varepsilon^2} = \int_{-\infty}^{\infty} \varphi_2(\omega)\, \mathrm{d}\omega = \int_{-\infty}^{\infty} |F_2(j\omega)|^2\, \varphi_N(\omega)\, \mathrm{d}\omega.$$

The requirement is to reduce to a minimum the difference $\varepsilon(t)$ between the measured value $s(t)$ and the value $s_0(t)$ known beforehand from the statistical properties of the signal. To do this, we attempt to minimize the mean square value of ε, $\overline{\varepsilon^2(t)}$, where $\varepsilon = s_D(t) - s(t)$.

When an interception occurs, ε represents the minimum trajectory between the missile and its target.

The optimization calculation takes into account the parameters describing the behaviour of the target. If, on the other hand, the target knows the transfer function of the missile and its guidance system, he can manoeuvre in such a way as to make the separation a maximum.

It is for this reason that a filter which appears to be the best when random accelerations, both in amplitude and direction, can occur may be inefficient for some special form of behaviour on the part of the target (level straight flight, for example), whereas a simple phase-advance filter would have been adequate.

There is a difference between calculations designed to obtain a filter in the most probable conditions, and those appropriate to the worst possible situation.

On the other hand if the missile were extremely effective against a target in straight flight, its effectiveness against a rapidly varying target, with sharp lateral acceleration, would be lost.

If an interception is to be ensured in every case, the missile should be capable of displaying "intelligence", of appreciating what kind of behaviour the target is displaying; it would then adapt its own parameters accordingly —which is known as "self-optimization".

VII.5.2. *Self-optimization*

Suppose that both the effective values of the variations of noise and those of the movements of the target vary, while the appearances of the spectra remain unaltered.

If we are to be able to regard the acceleration signal as statistically random, its spectral density must be of the form $(A/\omega^2)^2$, in which A is a constant and ω is the angular frequency in question. White noise corresponds to a spectral density of the form B^2 (B = constant).

Suppose the system is that of Fig. VII.30.

The difference between the output signal and the input signal onto which the noise is superimposed constitutes the apparent error ε. Here $\varepsilon = \dfrac{n + e}{1 + K}$ in which $K = G \dfrac{1 + pT}{p} \cdot \dfrac{1}{p}$.

The transfer function which relates the error to the input signal perturbed by noise is

$$\frac{\varepsilon}{n+e} = F = \frac{1}{1+K}.$$

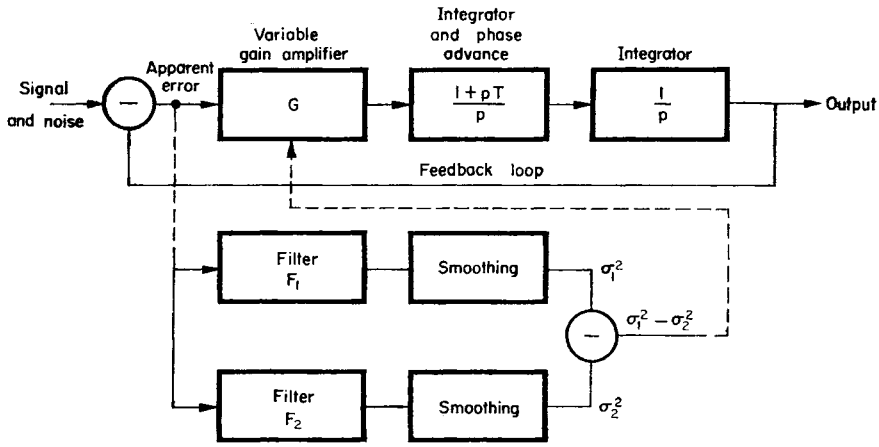


FIG. VII.30. Self-optimizing system.

The spectral density which corresponds to ε is therefore

$$\varphi_\varepsilon(\omega) = \frac{1}{|1+K|^2}(\varphi_N + \varphi_E);$$

but

$$K(p) = G\frac{1+pT}{p}\cdot\frac{1}{p} = \frac{G}{p^2}(1+pT),$$

and so

$$|1+K|^2 = \left|1 + \frac{G}{(j\omega)^2}(1+jT\omega)\right|^2,$$

or

$$|1+K|^2 = 1 - \frac{2G}{\omega^2} + \frac{G^2}{\omega^4} + \frac{G^2T^2}{\omega^2}.$$

Selecting $T = \sqrt{(2/G)}$, we find

$$|1+K|^2 = 1 + \frac{G^2}{\omega^4},$$

and the spectral density becomes

$$\varphi_\varepsilon(\omega) = \frac{B^2 + \dfrac{A^2}{\omega^4}}{1 + \dfrac{G^2}{\omega^4}} = B^2 \frac{\omega^4 + \dfrac{A^2}{B^2}}{\omega^4 + G^2},$$

in which G is the gain in the loop and $T = (2/G)^{1/2}$ is the time constant of the integrator.

It is to be noticed that by choosing $G = G_{opt} = A/B$, the spectral density is rendered independent of the frequency, and becomes equal to B^2. This is a criterion of quality for the system to be optimal in the presence of white noise.

The mean square error is then

$$\overline{\varepsilon^2} = \int_{-\infty}^{\infty} \varphi_\varepsilon(\omega)\, \mathrm{d}\omega$$

which is equal to $4A^{1/2}B^{3/2}$.

In order to adjust the gain G to the value A/B at every instant, we use as criterion the stability of the spectral density of the error. In fact, when G is greater than its optimum value, the spectral density increases with frequency, and vice versa. A comparison between the output signals of the two filters F_1 and F_2, therefore will indicate whether the gain G should be increased or decreased. These two filters cover different parts of the spectrum; if their bandwidths are equal, the gain G will be optimal when $\Gamma_1^2 - \Gamma_2^2 = 0$. The sign of $\Gamma_1^2 - \Gamma_2^2$ indicates the sense in which the the gain must be adjusted to attain the optimum condition.

We have assumed in the preceding discussion that the signal at the input and the noise are stationary. The power of the signal and the noise do in fact vary slowly, and to retain the conclusions reached above, the approximation that the behaviour is quasi-stationary during a certain time interval must be made. The expressions Γ_1^2 and Γ_2^2 are in fact obtained after smoothing the information.

VII.5.3. *Combined Filtering—Application to Mixed Guidance, using Velocity and Acceleration*

The problem is to measure some physical parameter by means of two independent techniques. If the spectral distributions are different, we can perform a filtering free of distortion.

A velocity can be measured either by differentiating the data regarding the position, or by integrating a measurement made upon the acceleration.

Where differentiation is concerned, the error is greater at higher frequencies and extremely small for frequencies close to zero; whereas when a signal has to be integrated, the variations of the envelope of the error spectrum are inversely proportional to the frequency.

Suppose that F_a and F_p are the transfer functions of the filters, which correspond to measurements of acceleration and of position respectively; N_a and N_p are the corresponding errors (see Fig. VII.31). V is the actual velocity, and V_D is the estimate of this velocity given by the circuit.

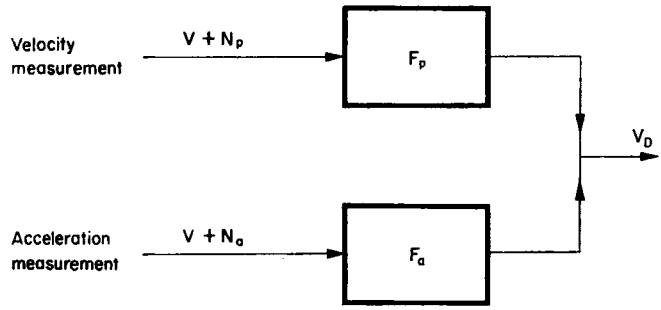


FIG. VII.31. Combined filtering.

The error in estimating V is thus

$$\varepsilon_V = V_D - V = F_p(V + N_p) + F_a(pV + N_a) - V$$

$$= (F_p + pF_a - 1)V + F_pN_p + F_aN_a.$$

The system is free of distortion if

$$F_p + pF_a - 1 = 0.$$

Assuming that the mean noise levels, N_p and N_a, are zero, and that these functions are stationary and random, the spectral density of the error in the velocity, ε_V, is then

$$\varphi_{\varepsilon V} = |F_p|^2 \varphi_p + |F_a|^2 \varphi_a.$$

Presuming the system to be free of distortion, we have

$$F_a = \frac{1 - F_p}{p},$$

so that

$$\varphi_{\varepsilon V} = F_p^2 \varphi_p + (1 - F_p)^2 \cdot \frac{\varphi_a}{\omega^2}$$

in which φ_p and φ_a are the spectral densities which correspond to N_p and N_a respectively. The mean square error in the velocity is of the form

$$\overline{\varepsilon_V^2} = \int_{-\infty}^{\infty} \varphi_{\varepsilon V} \, d\omega,$$

and it is therefore possible to find the expression for F_p which makes $\varphi_{\varepsilon V}$ a minimum if φ_p and φ_a are known.

The improvement to be obtained with combined filtering is a function of the degree of difference between the spectral densities of the two systems being measured. More than two filters can clearly be combined, and the accuracy cannot but be better. It is to be noticed that the theory of filtering can be successfully applied to a device with diversity.

More particularly, the preceding analysis can be applied to mixed navigational systems, using inertia and radio, in which the measurement of the acceleration of the inertial system and measurement of the speed by means of the Doppler effect are combined.

Suppose that the noises on the Doppler channel and on the accelerometer channel are uniform, and that their spectral densities are $\varphi_a = k_1^2$ and $\varphi_p = k_2^2$ respectively.

The quantity $\varphi_{\varepsilon V}$ reaches a minimum when $k_1 F_p = k_2 \dfrac{(1 - F_p)}{p}$, and we can thus deduce that the filter to be used must obey

$$F_p(p) = \frac{1}{1 + \dfrac{k_1}{k_2} p} = \frac{1}{1 + pT}$$

in which $T = k_1/k_2$.

From F_p, we deduce that the filter F_a should obey

$$F_a = \frac{1 - F_p}{p} = \frac{1 - \dfrac{1}{1 - pT}}{p},$$

or

$$F_a = \frac{T}{1 - pT}.$$

CHAPTER VIII

RADIO TELEMETRY EQUIPMENT

INTEREST in telemetry began to grow in the early 1930's, in particular in connection with meteorological work performed with balloons.

Nowadays, telemetry is widely used and is often indispensable.

In this chapter, the features which are peculiar to telemetric equipment are examined, together with the ways in which the information is reconstituted; in conclusion, the chapter is illustrated with a few actual examples.

VIII.1. GENERAL REMARKS

VIII.1.1. *The Field of Application*

Where missiles are concerned, two broad categories should be distinguished: research missiles, which may or may not be remote-controlled, and which carry complicated measuring apparatus for studying the upper atmosphere, for example, or even extraterrestrial space, and military missiles. In the first case, the role of telemetry is an obvious one, since recording equipment is difficult to recover intact. In the second case, this is no longer true, for if the missile in question is a remote-controlled target aircraft, it is possible to recover it after the flight, and all the essential parameters can be recorded by standard methods. The net cost of such devices is more modest than that of telemetry equipment, and further, the parameters are recorded directly, onto photographic paper for example.

It seems, in consequence, that we might hesitate before using a telemetry device if a simple recorder would suffice. However–even where target aircraft are concerned, this presents numerous inconvenient features.

When more than six parameters have to be recorded, it rapidly becomes difficult to distinguish the traces given that the recorders must be very compact.

It is above all important to be able to measure the parameters when the missile is still in the study stage. A test flight provides no useful information when the missile is destroyed. Research on missiles obliges one therefore to use radio telemetry, to know what is happening within the missile during a flight.

VIII.1.2. *Quantities which are Measured by Telemetry*

There are numerous measurements which can be made aboard a moving missile and which one would like to be transmitted through a radio link. Physical quantities which are harmful to the course of the flight are detected, and their influences noted. The flight parameters which, after interpretation, describe the aerodynamic behaviour of the missile, must be recorded. And finally, it is sometimes necessary to record the variations of some particular physical phenomenon.

Physical quantities which are harmful include vibrations, accelerations, increase of temperature, and internal pressures which result from distortion of the hull during the flight.

The flight parameters of interest are angular displacements and angular velocities about the axes of yaw, pitch and roll, the static and dynamic pressures, the fuel consumption and the speed of the missile.

Finally, the missile may have to measure some physical phenomena such as atmospheric radioactivity, ionization and humidity.

There is also the application of telemetry to the retransmission of a television image, either directly or after storing it in a memory.

All these quantities will be transformed into electrical quantities for radio transmission. The operation is effected by a special unit in the telemetry equipment itself, which will be studied in the second section of this chapter.

VII.1.3 *Criteria for a Telemetry Measurement*

The criteria associated with a measurement are fidelity, accuracy, and sensitivity. To these general characteristics must be added

- the need to transmit a great many measurements along a single link
- cross-modulation between telemetry channels must be reduced to the minimum
- the percentage distortion must be as low as possible
- an efficient protection against noise must be provided.

All these criteria can be simply united by defining the information capacity of the telemetry system and the information efficiency associated with each channel.

It is also worth emphasizing that materials carried in an aircraft – and more generally, those employed in telemetry – undergo physical conditions which are often severe (such conditions as those affecting temperature, pressure, acceleration, vibration and salinity). Finally everything on board must be as compact as possible, compatible with the transmitter power.

The problem of reconstructing the data will be treated later; however it should be pointed out at this stage that the easier it is to interpret the telemetry data rapidly on the ground, the more valuable is the telemetry equipment.

VIII.1.4 *FM–FM or FM–AM Telemetry*

In choosing the frequency and the distribution of the subcarrier waves, one must take into account the width of the desired pass-band for measuring possible cross-modulation, and the spacing which is necessary between the subcarrier waves if the channels are to be sufficiently separated.

By using frequency modulation of these subcarrier waves, defects in the fidelity of the elements of the transmission chain can be avoided.

FM–FM or FM–AM telemetry is preferable when one wants to transmit continuously over all the channels, and when the required pass-band width is large. Further, these arrangements are well-adapted for transducers with variable inductance or capacitance (although this is not so when variable resistance transducers are used).

The fidelity of FM–FM or FM–AM equipment is related to the frequency stability of the oscillator. To reduce the effects of drift, the frequency changes can be increased.

Under the best conditions, an accuracy of about 1 per cent can be obtained when calibration and balancing have been performed just before the mesaurement is made; however a figure of about 3 to 5 per cent is more likely when measurements last an appreciable time.

VIII.1.5 *Telemetry with Commutation*

In commutation equipment, the data from the different measurements are switched by means of a commutator which connects the various measuring detectors to the modulator in turn and sequentially. The commutator may be either mechanical or electronic.

At the receiver, a synchronized commutator separates out the different pulses in the channel which are then demodulated.

As the measurements are discontinuous, the bandwidth is bounded. The number of channels therefore depends upon the width of this pass-band. A mechanical commutator can give about 1000 contacts/second. With a pass-bandwidth of a few cycles/second, and with 30 contacts/second, about thirty telemetry channels are possible.

Voltages are easy to transmit in this manner, but transmission of a frequency is a more delicate task.

It is possible to attain high accuracy with this method by using a few of the channels in the sequence for transmitting fixed voltages which act as reference voltages at the receiver.

The two arrangements – continuous and with commutation, can be combined. We have then only to select one subcarrier wave which we commutate. The continuous channels are reserved for measurement which require a

large bandwidth. In this way a great many telemetry channels can be obtained. It should be noticed that there is a risk that mechanical commutation of induction detectors may lead to certain perturbations in transient phenomena.

VIII.2. FEATURES PECULIAR TO TELEMETRY: THE TRANSDUCERS

The function of transducers is to transform physical quantities into electrical quantities; the electric signal modulates the corresponding ratio channel. Transducers are briefly examined in this section.

Transducers are of five kinds:

- variable resistance (strain gauge)
- variable inductance
- variable capacitance
- variable resistance (potentiometer)
- variable voltage

In the following table, the uses to which the various kinds of transducer can be put are summarized.

TABLE VIII.1. TYPES OF TRANSDUCER

Quantity to be measured	Capacitative	Inductive	Magnetic	Piezoelectric	Photoelectric	Radioactive	Resistive	Thermo-electric
Acceleratoin	×	×	×	×			×	
Displacement	×	×	×	×	×	×	×	
Output	×	×	×	×		×	×	
Force		×		×		×	×	
Humidity and moistness	×						×	
Level	×			×	×	×	×	
Brightness					×		×	×
Mass		×	×	×		×		
Pressure	×	×	×	×		×	×	×
Temperature					×	×	×	×
Thickness	×	×		×	×	×		
Speed	×	×	×	×	×	×	×	
Viscosity	×			×			×	

VIII.2.1. *Strain-gauge*

These transducers are basically designed to measure strains. They consist of a resistive metal wire (see Fig. VIII.1) fixed to a thin sheet of paper.

The surface area $a \times b$ of the strain-gauge is very small, so that strains can be very easily localized. Variations of the resistance are due to variations of the length and hence of the cross-section of the wire.

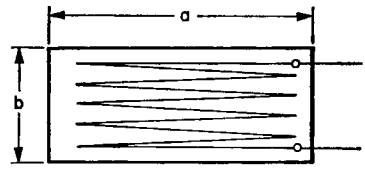


FIG. VIII.1. Resistance wire strain gauge.

Let $R = \varrho l/s$ be the resistance of such a gauge. If the wire is extended by an increment Δl, and l is the length of the resistive wire, we find

$$\frac{\Delta R}{R} = G \frac{\Delta l}{l}$$

in which G is the sensitivity factor or gauge factor; if η is Poisson's ratio, we have

$$\frac{\Delta R}{R} = (1 + 2\eta) \frac{\Delta l}{l}$$

and the sensitivity factor becomes

$$G = 1 + 2\eta + \frac{\varrho_0 - \varrho}{\varrho_0}$$

where ϱ_0 is resistivity of the unstretched wire .

For accurate measurements, the effect of temperature is taken into account. If t is the temperature coefficient of the material in question, the resistance at t°C is given by

$$R = R_0[1 + \alpha(t - t_0)].$$

The relative change of resistance due to this temperature effect is therefore

$$\frac{\Delta R}{R} = \alpha(t - t_0),$$

and this is equivalent to a fictitious strain given by

$$\alpha(t - t_0) = G \frac{\Delta l}{l}.$$

Hooke's law gives Δl as a function of the applied stress:
$\Delta l = Fl/ES$, where E is Young's modulus. The resulting error in the stress measurement is thus:

$$\frac{F}{S} = \frac{E\alpha(t - t_0)}{G}.$$

To avoid temperature effects, however, two or sometimes four identical strain gauges are placed in a Wheatstone bridge circuit.

In the following table, values are given of the sensitivity factor G and of the temperature coefficient for various kinds of substance.

TABLE VIII.2[52]

Material	Gauge factor	Temperature coefficient α ohm/ohm°C
Chromel	2·5	—
Constantan	2	0·00001
Manganese	0·47	0·00001
Nichrome	2·5	0·0004
Nickel	−12·1	0·006
Phosphor bronze	1·9	0·002
Platinum	6	0·003

The arrangement shown in Fig. VIII.2 employs four strain gauges in a bridge circuit. The output voltage is given by

$$E_s = e\frac{R_1R_4 - R_2R_3}{(R_1 + R_2)(R_3 + R_4)}.$$

At rest, the balance condition $R_1R_4 = R_2R_3$ is satisfied, so that $E_s = 0$.

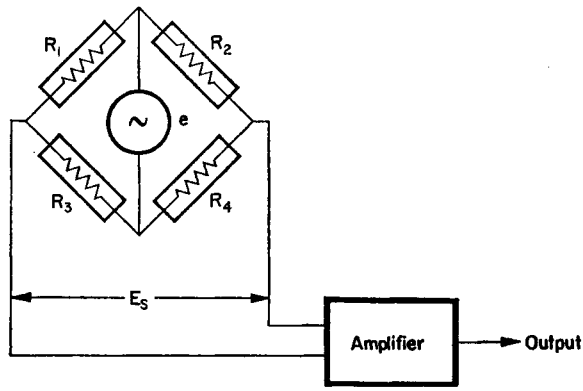


FIG. VIII.2. Bridge circuit.

When the strain gauge resistances vary by $\Delta R_1, \Delta R_2, \Delta R_3$ and ΔR_4, the output voltage becomes

$$E_s = e\frac{R_1\Delta R_4 + R_4\Delta R_1 - R_2\Delta R_3 - R_3\Delta R_2}{(R_1 + R_2)(R_3 + R_4)}.$$

Usually we arrange that $R_1 = R_2 = R_3 = R_4 = R$, say, to ensure maximum sensitivity, and hence

$$E_s = e\frac{(\Delta R_1 + \Delta R_4) - (\Delta R_2 + \Delta R_3)}{4R}.$$

This out of balance voltage is amplified and used to modulate the subcarrier waves; the LF frequency is that of e.

VIII.2.2. *Variable Voltage or Variable Current Transducers*

These elements can be further subdivided into piezoelectric, photoelectric, thermoelectric, magnetoelectric, electrochemical, electronic and radioactive elements. We shall consider only the first two types.

VIII.2.2.1. *Piezoelectric Transducers*

When a pressure is applied to a piezoelectric crystal, a potential difference appears across its faces which is proportional to the applied stress. This phenomenon occurs with quartz, for example. The electric charge collected between the two faces across which the force F acts.

$$Q = K|\mathbf{F}|$$

in which K is the Curie constant ($K = 6{\cdot}45 \times 10^{-8}$ e.s.u.). This property is found for example in Rochelle salt and barium titanate. Synthetic crystals are more sensitive (1000 times) than quartz, but they have the inconvenience of being hygroscopic.

VIII.2.2.2. *Photoelectric Transducers*

An element which is sensitive to light flux—a photodiode or a phototransistor, for example—delivers a current which is proportional to the incident illumination density. Good sensitivity and a rapid response-time can be obtained from such a device.

The photoelectric effect can be used directly to measure such quantities as the absorption of a gas, the influence of reflections, the level of a liquid, or the solidification of a body. As an example, a diagram showing a photoelectric device used for the measurement of pressure is given.

A very thin, circular, elastic diaphragm is mounted under tension as shown in Fig. VIII.3. The radius of curvature of the diaphragm is inversely proportional to the applied pressure difference, $(p_1 - p_2)$. If the surface of the diaphragm is reflecting, an incident parallel light beam is reflected as either a divergent or a convergent beam, depending on the sign of $(p_1 - p_2)$. A

photoelectric cell, relatively close to being a point, will therefore receive a variable light intensity.

Silver foil can be used for the diaphragm and the device can be made linear by arranging it in the symmetrical form shown in Fig. VIII.4.

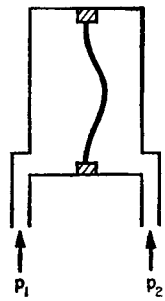


FIG. VIII.3. Measurement of pressure by change in curvature of a diaphragm.

The response-time depends upon the dimensions of the cavity of the diaphragm; natural frequencies are about 500 c/s.

The pressure gradient creates variations in the radius of curvature, and this is related to the thickness, the tension and the elasticity of the diaphragm.

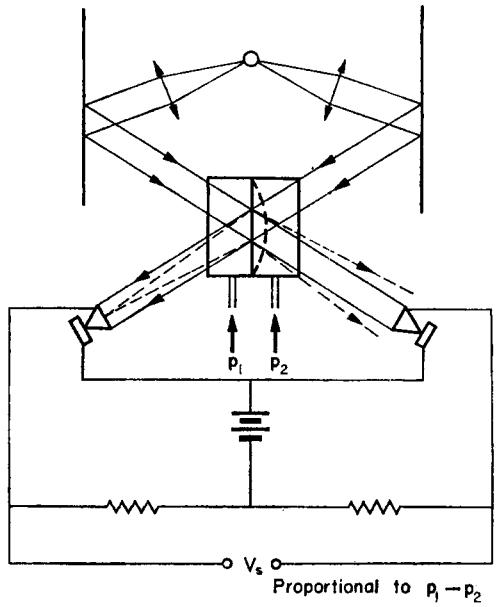


FIG. VIII.4. Balanced photoelectric pressure transducer.

VIII.2.3. *Transducers with Variable Inductance*

This is the kind of transducer found most frequently in telemetry equipment. Its output signal is directly usable.

Consider a winding with n turns, through which a current I flows; the e.m.f. appearing across the terminals of this circuit is

$$e = -n\frac{d\varphi}{dt} \times 10^{-8}\ \text{V}.$$

The flux is related to the current through the inductance L, ($\varphi = LI$), which is in turn given by $L = \dfrac{4\pi N^2}{\mathscr{R}} \times 10^{-9}$ henries; $\mathscr{R}$ is the reluctance of the magnetic circuit:

$$\mathscr{R} = \sum \frac{1}{\mu}\frac{l}{s}.$$

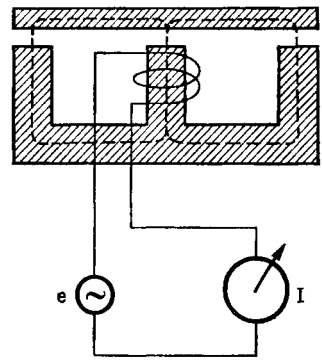


FIG. VIII.5. Magnetic circuit with variable reluctance.

Consider the arrangement shown in Fig. VIII.5. We have $\mathscr{R} = \dfrac{l}{\mu s} + \dfrac{x}{s}$, in which x is the air-gap (rest value, x_0). l is the mean path length in iron and s is the area of cross-section.

If l/μ is negligible in comparison with x, then

$$\mathscr{R} \simeq \frac{x}{s}$$

and

$$L = \frac{4\pi N^2 S}{x}.$$

The relative changes in inductance and gap-width are related by

$$\frac{\Delta L}{L} = -\frac{\Delta x}{x_0},$$

and the current flowing in the coil is

$$i = j\frac{e}{\omega L}.$$

The current is thus modulated by the variations in the inductance. If the coil forms part of the tuned circuit of an oscillator, the variation of self-inductance will be directly retransmitted as a variation of frequency. For example, an absolute manometer, built by the SFIM, is shown in Fig. VIII.6.

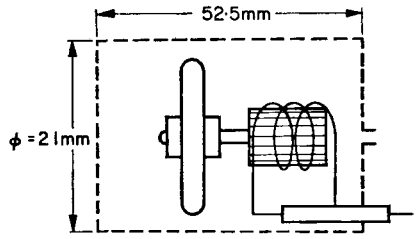


FIG. VIII.6. Absolute manometer with variable inductance.

This manometer measures pressures in the ranges 800 mb, 1300 mb, 2500 mb, 4000 mb or 10,000 mb. The zero of the device can be set at any one of these values.

The hysteresis error is smaller than 0·5 per cent. For a temperature variation of $-30°$ to $+70°$, the error remains less than 2 per cent. It should be mentioned that this device is affected by accelerations parallel to the measuring axis.

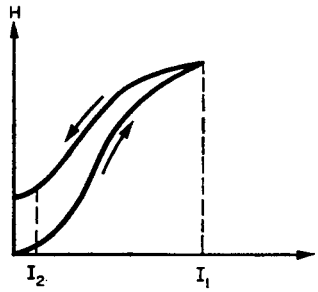


FIG. VIII.7. The hysteresis phenomenon.

Generally speaking, the accuracy of a variable-permeability transducer is limited by temperature and the effect of acceleration.

Because of hysteresis phenomena, the current must have attained its maximum value at least once before every measurement and, during the measurements, it should never exceed the limits I_1, I_2 (see Fig. VIII.7).

During a temperature cycle, it can be seen that the frequency variation Δf drifts about the rest value. The maximum frequency departure Δf, as a function of temperature (see Fig. VIII.8), shows the resulting error.

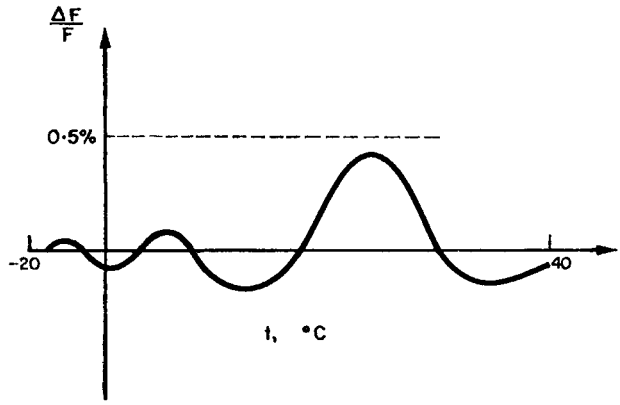


FIG. VIII.8. The relative variation of the oscillation frequency as a function of the temperature.

VIII.2.4. *Transducers with Variable Capacitance*

The capacitance of a parallel-plate capacitor is

$$C = \frac{\varepsilon_r S}{4e}$$

in which S is the surface area of the plates and e is the thickness of the dielectric (ε_r) between them.

Differentiating C with respect to e gives

$$dC = -\frac{\varepsilon_r S}{4\pi} \frac{de}{e^2}.$$

Changes in capacity are not linearly related to the distance between the two plates.

To measure a pressure, a capsule composed of two plates, one fixed and the other a metallic membrane, can be employed (see Fig. VIII.9). The deformation of the membrane is parabolic, and the variation law is

$$e = e_o + \Delta e \left(1 - \frac{r^2}{R^2}\right).$$

The total capacitance then becomes

$$C = \int_0^R \frac{2\pi r \, dr}{4\pi \left[e_0 + \Delta e \left(1 - \frac{r^2}{R^2} \right) \right]}$$

$$= \frac{S}{4\pi e_0} \left[1 - \frac{\Delta e}{2e_0} + \frac{(\Delta e)^2}{3e_0^2} \cdots \right].$$

To a first approximation

$$\frac{\Delta C}{C} \simeq - \frac{\Delta e}{e_0} \left[\frac{1}{2} - \frac{\Delta e}{3e_0} \right].$$

If the deformable part of the capsule consists of a circular membrane clamped round its circumference, the central deflection of the membrane, as a function of the uniformly distributed load p, is equal to

$$e = \frac{3}{16} (1 - \Gamma^2) \frac{p}{E} \frac{R^4}{8s^3}$$

in which Γ is Poisson's ratio, R the radius, s the thickness of the membrane and E Young's modulus.

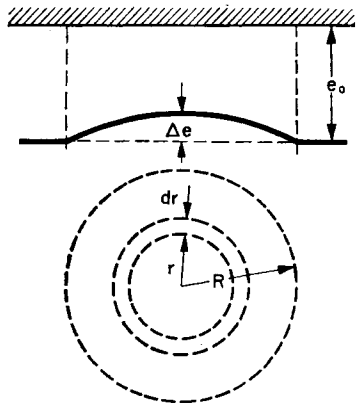


FIG. VIII.9. A capsule with a membrane.

Thus Δe is proportional to p. If $\Delta e/e_0$ is small enough for the second order term to be negligible, the relative variation of the capacitance is effectively equal to

$$\frac{\Delta C}{C_0} \simeq \frac{\Delta e}{2e_0}$$

and is thus proportional to p.

When the capacitor C is inserted into an oscillatory circuit, with $LC\omega^2 = 1$ we have

$$\frac{\Delta C}{C} = -2\frac{\Delta\omega}{\omega},$$

or

$$\frac{\Delta\omega}{\omega_o} \simeq \frac{1}{4}\frac{\Delta e}{e_o}.$$

To attain a linearity of x per cent in frequency modulation by variation of the capacity in an oscillatory circuit, the inequality

$$\frac{\Delta C}{C_o} \leqslant \frac{8}{3}\frac{x}{100}$$

must be satisfied. Thus a compromise can be made between the sensitivity of the capsule and the desired linearity.

Nowadays, technical progress is such that variable capacitance can be obtained by the use of silicon diodes; the "varicaps", mentioned in the preceding chapter, provide an example of this. The capacity of such a diode varies with the bias applied across the junction (which must be biased in the reverse sense). These diodes can give variations from 30 to 3 pF for bias voltages continuously varying between 0·2 and 50 V. This results in a variable capacitance transducer of extremely small volume.

VIII.2.5. *Transducers Employing Potentiometers*

The mechanical displacement of an object can be transmitted to the wiper of a potentiometer; a direct or alternating voltage is thus obtained, which is a function of the position of the wiper.

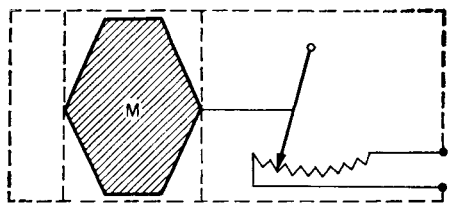


FIG. VIII.10. The potentiometer accelerometer J 4201

Such a device for measuring acceleration is shown in Fig. VIII.10, the SFIM model J 42. The mass M is suspended elastically, and transmits a measure of the acceleration to the wiper. The resistance of the potentiometer is 1000 Ω and there are 250 turns.

Depending upon the type of accelerometer employed, the range from $\pm 0{\cdot}6g$ to $\pm 12g$ (in which g is the acceleration due to gravity) can be measured. The model J 4201 has a range of measurement of $\pm 0{\cdot}6g$ and its sensitivity is $0{\cdot}12g$ for 100 Ω. The e ror in the movement of the wiper is ±0·5 per cent, and for a temperature variation of −30 °C to +70 °C it becomes 4 per cent.

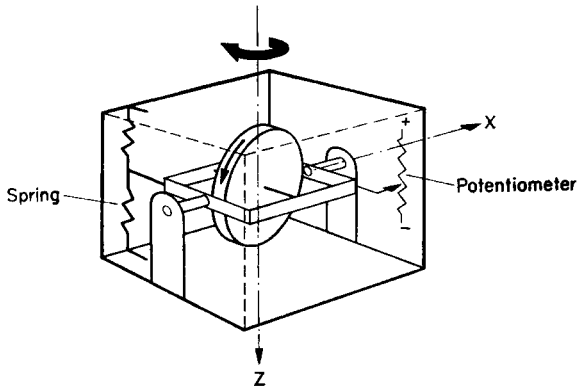


FIG. VIII.11. Rate gyroscope.

The natural frequency of such a device is given by the relation $f = 13\sqrt{n} \pm 0{\cdot}15$ in which n is the total number of g. For the J 4201 accelerometer, for example, $f = 14$ c/s.

Another interesting example of a transducer is the rate gyroscope. For example the I 14 rate gyro of the SFIM, which is used for measuring angular velocities is shown in Fig. VIII.11. The spinner is driven by a D.C. motor and its speed regulated by a centrifugal switch. The gyroscopic torque, proportional to input rate, produces a proportional deflection of the restraining spring which is in turn transmitted to the wiper of a potentiometer.

VIII.3. NON-LINEARITY IN TRANSDUCERS[34]

Let x and y be the magnitudes at the input and output of the transducer; between these two signals, a relation $y = f(x)$ exists as shown plotted in Fig. VIII.12. The line (*a*) represents the ideal curve; the line (*c*) passes through the point (x_M, y_M), which corresponds to the maximum measurement to be made; while the line (*b*) is an arbitrary line passing through x_0.

The departure from linearity, $\Delta y = g(x)$, represents the difference in ordinate between the function $y = f(x)$ and the lines (*a*), (*b*) or (*c*).

Expanding the function y in series form, gives

$$y = ax + bx^2 + cx^3 + \cdots$$

so that to a first approximation,

$$y = x - kx^2.$$

Let us regard the line (*b*) as the ideal curve for the transducer,

$$y_0 = x_0 - kx_0^2.$$

A point on the line (*b*) is characterized by

$$y' = \frac{y_0}{x_0} x = (x_0 - kx_0^2) \frac{x}{x_0} = (1 - kx_0)x.$$

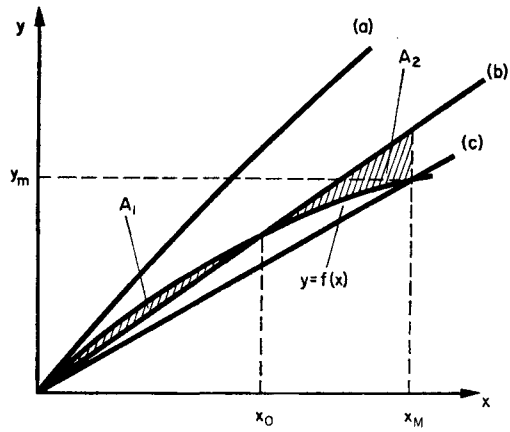


FIG. VIII. 12. The transducer characteristic.

The value of the disparity is

$$\Delta y = y - y' = k(x_0 x - x^2)$$

and this equation provides the non-linearity error.

We can discuss the value of the disparity in terms of the position of the point x_0. For $x_0 = 0$, $\Delta y = -kx^2$.

The maximum disparity is therefore $(\Delta y)_{max} = -kx_M^2$, which means that the measurement will depart farthest from the ideal when it is zero at the origin.

Choosing $x_0 = x_M$, the maximum departure due to non-linearity becomes

$$(\Delta y)_{max} = 0{\cdot}5kx_M^2$$

Differentiation of the equation $\Delta y = k(x_0 x - x^2)$ gives

$$\frac{d(\Delta x)}{dx} = k(x_0 - 2x),$$

and the departure is maximum when $x = x_0/2$, at which point it is equal to

$$[\Delta y]_{x=x_0/2} = \tfrac{1}{2}kx_0^2.$$

For $x = x_M$, Δy becomes

$$[\Delta y]_{x=xM} = k(x_0x_M - x_M^2).$$

Selecting x_0 in such a way that

$$[\Delta y]_{x=xM} = -[\Delta y]_{x=\frac{1}{2}x_0}$$

we find $x_0 = (\sqrt{3} - 1)x_M = 0{\cdot}732x_M$, and in this case the departure becomes

$$\Delta y = k(0{\cdot}732x_Mx - x^2).$$

This has a maximum when $x = x_M$, equal to

$$[(\Delta y)_{\max}]_{x=x_M} = -0{\cdot}368kx_m^2.$$

If the reference point is selected in such a way that the shaded areas A_1 and A_2 of Fig. VIII.12 are equal, we have

$$A_1 = k\int_0^{x_0} (x_0x - x^2)\,dx = \tfrac{1}{6}kx_0^3$$

$$A_2 = k\int_{x_M}^{x_0} (x_0x - x^2)\,dx = \tfrac{1}{6}kx_0^3 + \tfrac{1}{3}kx_M^3 - \tfrac{1}{2}kx_0x_M^2.$$

Putting A_1 equal to A_2, gives $x_0 = \frac{2}{3}x_M$ and substituting this into Δy gives

$$\Delta y = k[0{\cdot}667x_Mx - x^2].$$

For $x = x_M$, the departure is a maximum, or

$$[(\Delta y)_{\max}]_{A_1=A_2} = -0{\cdot}333kx_M^2$$

and this last is the more useful result.

VIII.4. THE CONVERTERS

Examples have been given of how mechanical quantities are transformed into electrical quantities by means of transducers. It is necessary to modify these electrical signals to make them suitable for radio equipment.

In certain cases, the transducer gives a quantity which can be used directly: variable inductance or capacitance transducers for example. These are then inserted into an oscillatory circuit. This is not always the case, however, and

sometimes a variation in a voltage, for example, has to be changed into a related variation of frequency; again, for transmission in coded modulation the voltage may have to be passed through an analogue–digital converter (this subject is treated in Chapter III—"Information and Coding").

VIII.4.1. *Direct Use of Variations of Inductance or Capacitance*

To use variations of these quantities directly, it is only necessary to insert the transducer into the circuit of an oscillator.

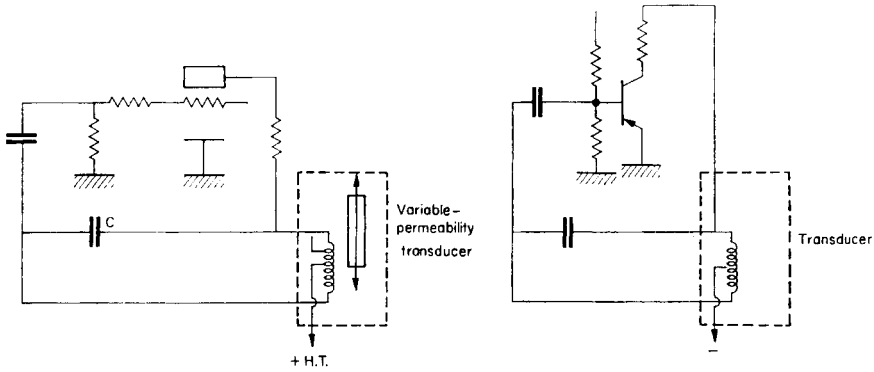


FIG. VIII.13. Hartley oscillators.

Consider the circuit in Fig. VIII.13. The Hartley oscillator oscillates at a frequency

$$f = \frac{1}{2\pi\sqrt{LC}}\left[1 - \frac{\mu' R}{2R_a' + (\mu' + 1)R}\right]^{1/2},$$

in which

$$\mu' = \frac{\mu R_c}{\varrho + R_c}, \quad R_a' = \frac{\varrho R_c}{\varrho + R_c},$$

and R = resistance of the inductor L
ϱ = valve anode slope resistance
R_c = anode load resistance
μ = amplification factor of the valve
C = tuning capacitance.

The condition for the oscillation to be maintained is

$$\frac{L}{C} = \frac{R[2R_a' + (\mu' + 1)R]^2}{\mu'[4R_a' + (\mu' + 1)R]}.$$

If the resistance of the coil is very small, and tends towards zero, the frequency of oscillation becomes $f = 1/2\pi\sqrt{LC}$ and the condition for oscillation vanishes.

A variation in the permeability modifies the inductance and results in a change of the oscillation frequency f.

VIII.4.2. *The Voltage Frequency Converter*

A variable voltage is to be converted into a variable frequency. The general arrangement is shown in Fig. VIII.14, and a detailed circuit is given in Fig. VIII.15.

The common emitter stage $T2$, introduces a phase change of $T1$ whereas the common collector stages $T1$ and $T3$ introduce no phase-change. The result is that if the circuit is to maintain oscillation, the feedback network must introduce a phase change of π and the loop gain must be made unity by $T2$.

By varying the input voltage, the output impedance of the transistor $T1$ is modified, and the oscillation frequency is shifted to ensure a total phase change of 2π. Good linearity is obtained for departures of ± 10 per cent on either side of the central frequency.

The same principle can be applied using a multivibrator (see Fig. VIII.16).

Consider now the arrangement with feedback shown in Fig. VIII.17, where $G = i_2/e_1$ and $\omega_0 = 1/\sqrt{LC}$ is the resonant frequency of the LC circuit.

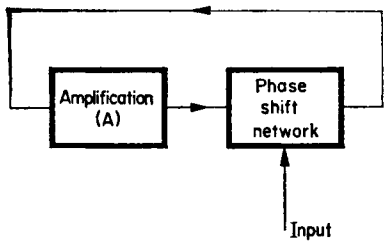


FIG. VIII.14. Voltage-variable frequency oscillator loop.

If the voltage across the resistor R_L is e_{RL}, the ratio M is defined by $M = e_M/e_{RL}$. In addition

$$K = \frac{e_5}{e_4}; \quad A = \frac{e_4}{e_3}; \quad Z_0 = \sqrt{\frac{L}{C}}; \quad Q = \frac{\omega_0 L}{R_L},$$

and thus

$$\frac{\omega}{\omega_0} = \sqrt{(1 - MR_L G - AKGR_L)}.$$

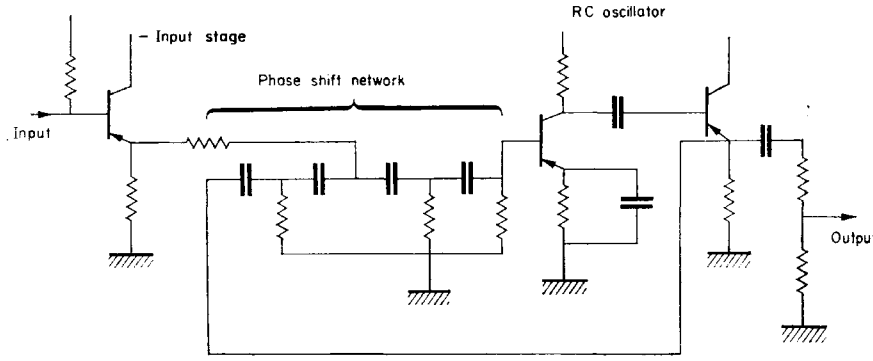


FIG. VIII.15. Voltage–variable frequency BC oscillator.

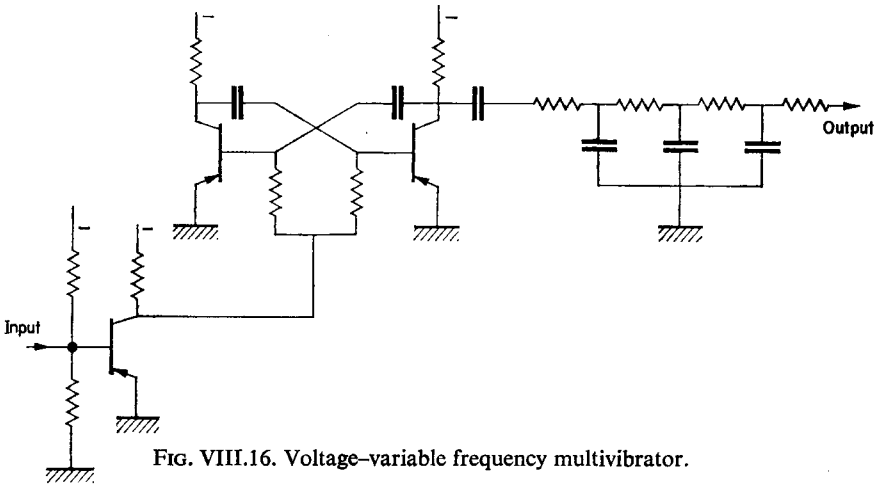


FIG. VIII.16. Voltage–variable frequency multivibrator.

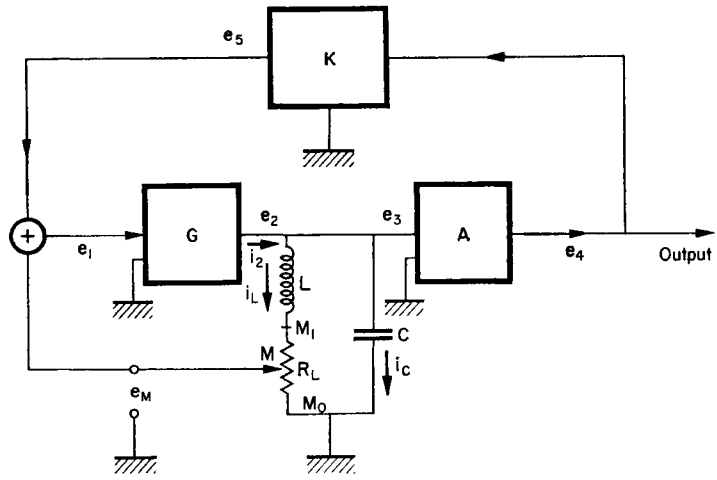


FIG. VIII.17. Linear oscillator.[91]

However,

$$AKG = R_L \cdot \frac{C}{L} = \frac{R_L}{Z_0^2} = \frac{1}{QZ_0},$$

and

$$R_L = \frac{Z_0}{Q} = \frac{\sqrt{L/C}}{Q},$$

so that

$$\frac{\omega}{\omega_0} = \sqrt{\left(1 - \frac{1}{Q^2} - MR_LG\right)}.$$

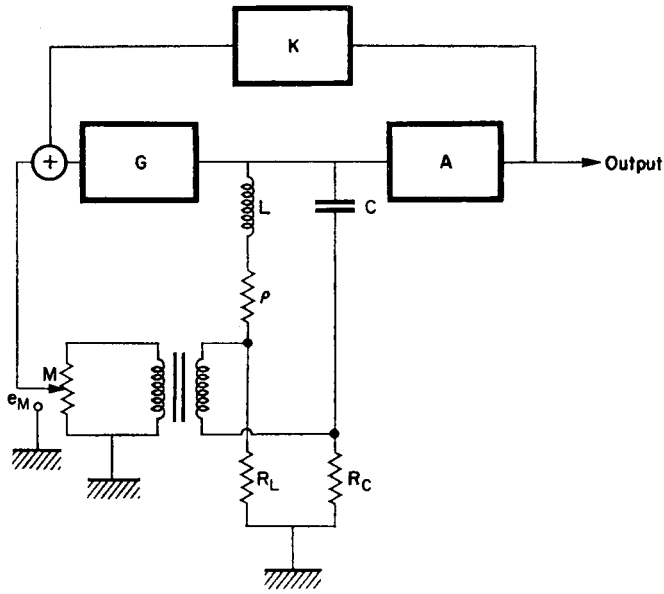


FIG. VIII.18. Linear subcarrier wave oscillator.

The quality factor Q is large enough for the term $1/Q^2$ to be negligible, and so

$$\frac{\omega}{\omega_0} \simeq \sqrt{(1 - MR_LG)}.$$

The variable parameter is M; it may represent for example the potentiometer of a transducer. A modified and improved version of the converter is shown in Fig. VIII.18.

The variation of ω/ω_0 when M varies from -1 to $+1$, is shown in Fig. VIII.19.

In this case the value of ω/ω_0 is equal to

$$\frac{\omega}{\omega_0} \cong \sqrt{\left[\frac{1 - MR_LG}{1 + MR_CG}\right]}.$$

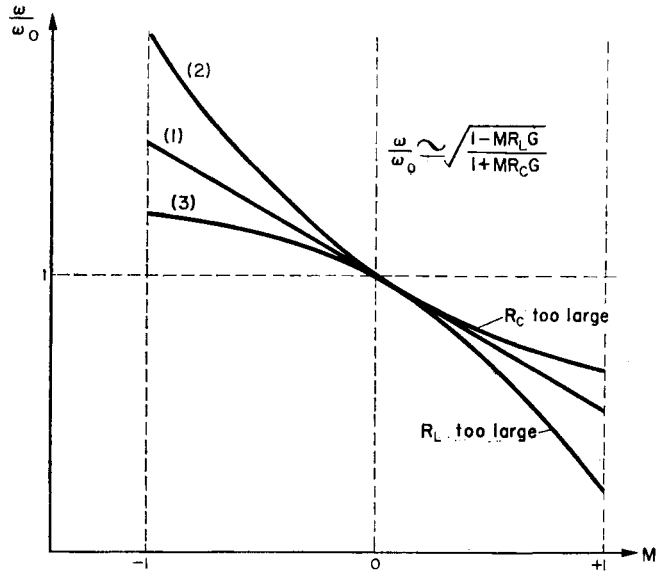


FIG. VIII.19. Variation of ω/ω_0 as a function of M.

The optimum value of $\frac{R_L}{R_C} \simeq 3$ with $\frac{\sqrt{L/C}}{R_L + R_C} \geqslant 20$, in which case the characteristic (1) of Fig. VIII.18 is obtained. If R_C is too large, the char-acteristic becomes that of (2) and if R_L is too big, ω/ω_0 is represented by the curve (3).

VIII.4.3. *Commutation of Telemetry Channels*

Data can be switched by mechanical or electronic commutation.

Suppose it is desired to retransmit three measurements A, B, C, which have been provided by three transducers, on the same subcarrier wave; the channels can be commutated as shown in Fig. VIII.20.

The mechanical commutator is driven by a motor which successively connects the data from the three transducers to a subcarrier wave oscillator.

The commutator shown in Fig. VIII.20 has 19 contacts: two of these are reserved for synchronization (Nos. 16 and 17), contact 18 provides a zero reference for the minimum voltage signal-level, and contact 19 corresponds to the maximum voltage which might be met in the course of a measurement.

The range of measurement is therefore given by $V_{max}-V_0$. Of the 19 contacts, 15 remain available for the three measurements A, B, C. Five contacts are hence allotted to each measurement:

contacts 1, 4, 7, 10, 13 for the measurement of A

contacts 2, 5, 8, 11, 14 for the measurement of B

contacts 3, 6, 9, 12, 15 for the measurement of C.

The resulting sampled signal on a channel is as shown in Fig. VIII.21.

By means of such a device, the number of telemetry channels can be multiplied. Consider the case of an FM/FM telemetry measurement where q channels are commutated on each subcarrier wave (see Fig. VIII.22).

If p subcarrier waves are available, and if each is commutated over q transducers, the total number of telemetry channels is $n = pq$.

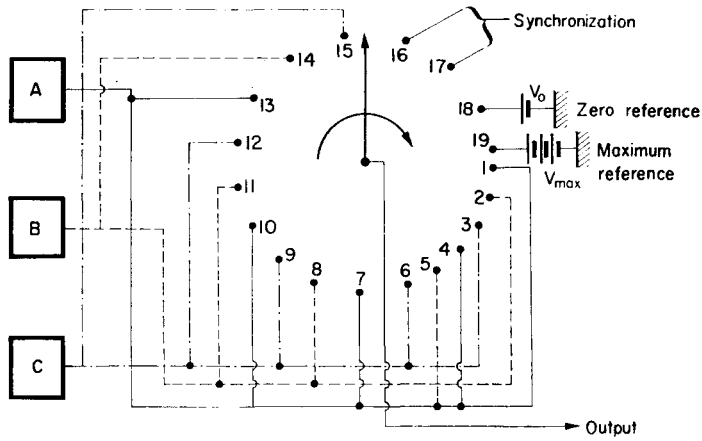


FIG. VIII.20. Channel commutation device.

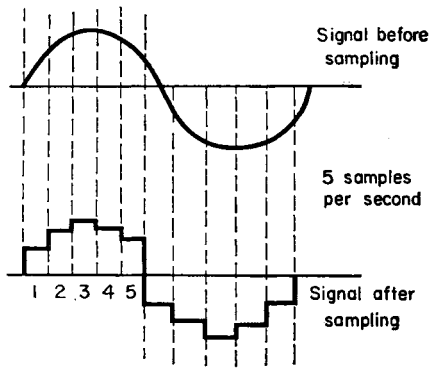


FIG. VIII.21. The sampled signal.

This is a most valuable procedure in telemetry equipment, as the bandwidth necessary for a measurement is usually of the order of a few cycles; measurements of vibration and of acceleration are exceptions.

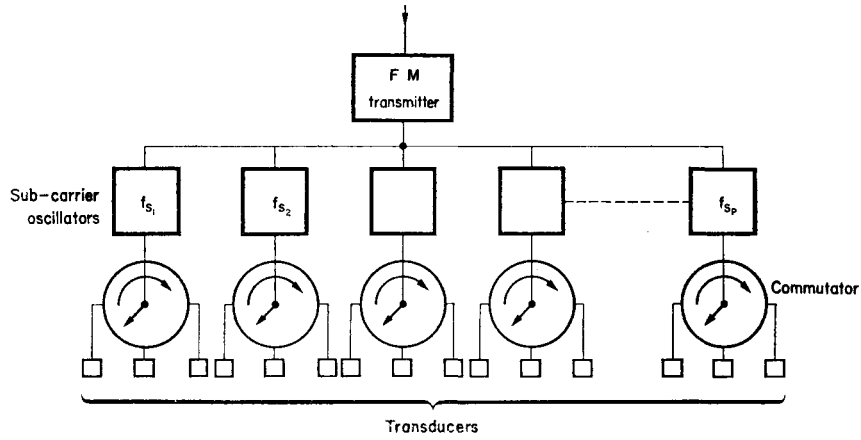


FIG. VIII.22. PAM–FM–FM system.

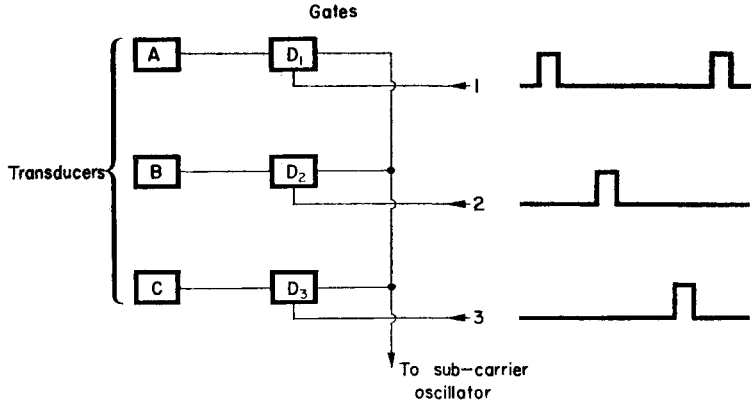


FIG. VIII.23. Electronic commutator.

Suppose that each of the measurements in question requires a bandwidth of 5 c/s. The sampling frequency will therefore have to be at least ten contacts per second. If three transducers are commutated onto the same subcarrier wave, the modulation bandwidth of each will have to be equal to at least 15 c/s.

The use of a mechanical commutator is less attractive when large bandwidths have to be transmitted, or a large number of channels commutated.

The speed at which the commutator can rotate is limited, and the contacts wear more rapidly at high speeds. In such cases, it is better to use electronic commutation. To do this, the three transducers A, B, and C are connected separately to three gates (D_1, D_2 and D_3). These only give an output when a pulse is applied (see Fig. VIII.23). It is arranged that the pulses (1), (2) and (3) are staggered in time with respect to one another, and the function of commutation is thus carried out.

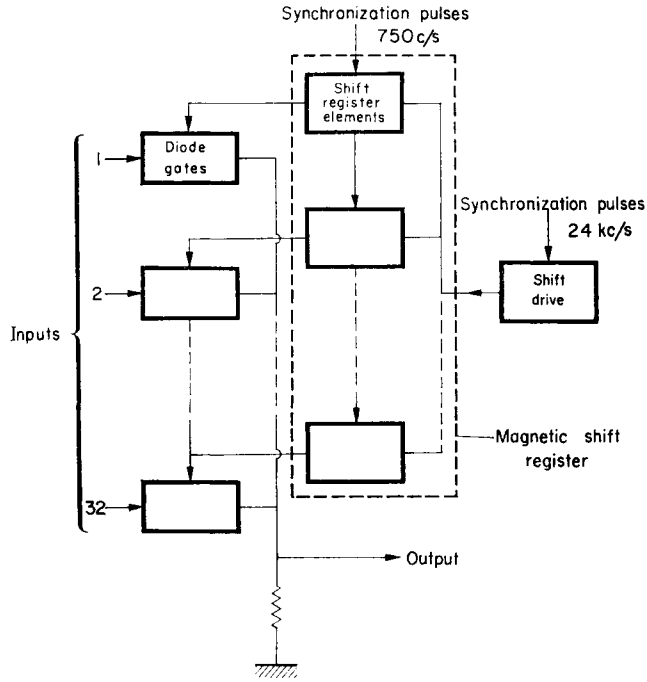


FIG. VIII.24. Electronic multiplexer.

The devices D_1, D_2 and D_3, can take the form of high speed relays such as UGON relays, or alternatively, purely electronic units, such as diode gates can be used.

A clock oscillator drives a binary ring-counter which in turn drives diode matrices. The diode gates are opened or closed every time a pulse from the clock oscillator passes through the counter.

By way of example, consider the electronic commutator or telemetry system multiplexer AN/AKT – $14^{(9,84)}$ in Fig. VIII.24. In 1/750th of a second the 32 channels are sampled, each for 1/24,000th of a second. The pulses at a frequency of 24 kc/s make the input pulse of the magnetic shift register advance step by step. At each shift, each of the 32 diode gates provides a pulse at the output, the level of which is proportional to the instantaneous value sampled.

The whole unit, comprising diode gates and magnetic shift register, is shown in detail in Fig. VIII.25.

The pulse at the output from the fourth winding produces a current I in each arm of the bridge. This current charges up a bias capacity. When the pulses are not present the diodes are cut off until the subsequent pulse causes them to conduct again. The forward and backward resistances of the diodes are 500 Ω and 100 MΩ, respectively, in series with the signal.

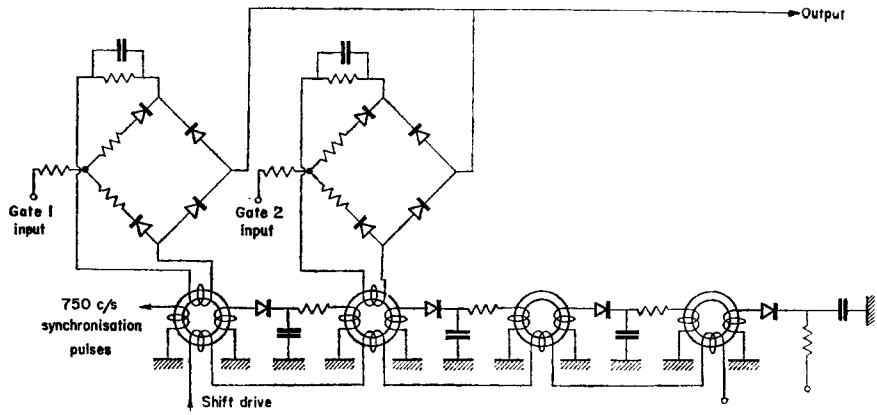


FIG. VIII.25. The "gates" and magnetic shift register.

VIII.4.4. *Decommutation at the Receiver*

At the receiver, the channels have to be decommutated: the various data have to be separated and fed into the storage device associated with each channel which reconstructs the data. To do this, first the synchronization and the sampling frequencies are detected; the first can be distinguished since at the transmitter the two contacts 16 and 17 of Fig. VIII.20 are short-circuited, so that the period associated with this synchronization is twice that of each individual contact. By forming saw-teeth from the rising and falling edges, the synchronization pulse is detected by limiting and differentiating, as shown in Fig. VIII.26.

The sampling frequency is obtained by differentiating either the rising or the falling edges.

Thus starting from the synchronization pulses we have a means of distributing the sampling levels sequentially in parallel form, in order to separate the telemetry channels.

An electronic decommutator with storage is shown diagrammatically in Fig. VIII.27.

The output from the decommutator is applied to gate circuits which distribute the appropriate data to each channel. After each gate is a storage unit, which may for example be a capacitor with a very slight loss, which plays the role of a memory. This memory retains the level of the pulse which has just been applied to it until the arrival of the next pulse. In short, the capacitor is charged up to the same level as the pulse.

To decommutate telemetry channels, a disc, identical to the mechanical commutator at the transmitter can equally well be used. This decommutator is incorporated in a control system in which the error voltage between the detected synchronization pulse and the corresponding pulse on the decommutator disc at the receiver is driven to zero. The error voltage modifies the speed at which the decommutator rotates, in order to bring the channel decoding into synchronization with the coding at the transmitter.

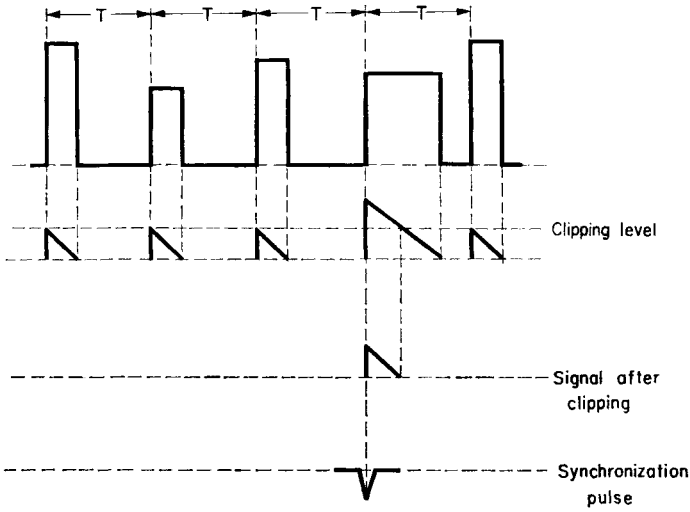


FIG. VIII.26. Detection of the synchronization pulse at the receiver.

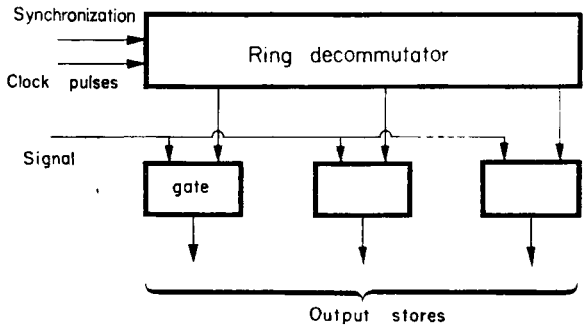


FIG. VIII.27. Electronic decommutator with storage.

VIII.5. SOME PRACTICAL EXAMPLES

VIII.5.1. *A PDM–FM Telemetry Transmitter*

After commutation, the variations of voltage are transformed into variations of pulse width; the transmission is carried out by PDM–FM.

The block diagram is given in Fig. VIII.28. A mechanical commutator supplies a synchronization signal, and the information from the transducers in sequential form. This information modulates the pulse width of a monostable multivibrator. The PDM modulated pulse drives a reactance valve, and provides a PDM–FM wave at the output. After multiplication and ampli-

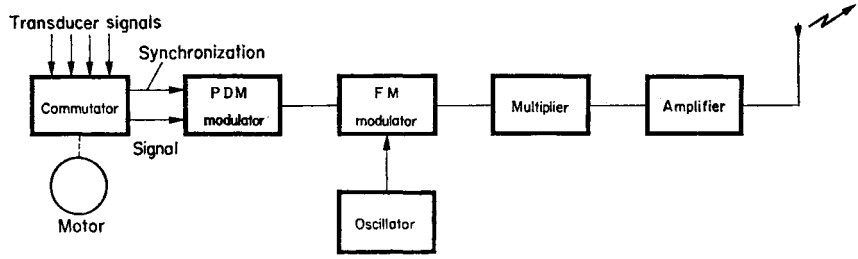


FIG. VIII.28. Block diagram of a PDM–FM remote-control transmitter.

fication, the modulated signal at the aerial is at about 4 watts and a frequency of 215 to 235 Mc/s. The speed at which the commutator motor rotates is 1200 r.p.m. and the number of commutator contacts is 45.

In Fig. VIII.29, a detailed circuit of the telemetry transmitter is shown.

VIII.5.2. *Description of a PCM–FM Telemetry Device (AN–AKT/14)*

This telemetry device has 32 channels which are sampled after amplification. The data is then in PAM form, and the telemetry channels appear in sequential form. A coder transforms the amplitudes of the various pulses into digital form and these digital pulses modulate a standard FM transmitter. In this way a PAM–PCM–FM arrangement is obtained (see the diagram in Fig. VIII.30).

The transducers give electrical signals of the order of 5 mV; the latter are sampled and applied to a.c. amplifiers.

The samples of information are made at a rate of 24,000/sec, and are coded into an eight digit binary system. The individual pulses of the code group last 4·7 μs, the deviation from the HF carrier wave is ±65 kc/s and the minimum HF bandwidth is 200 kc/s.

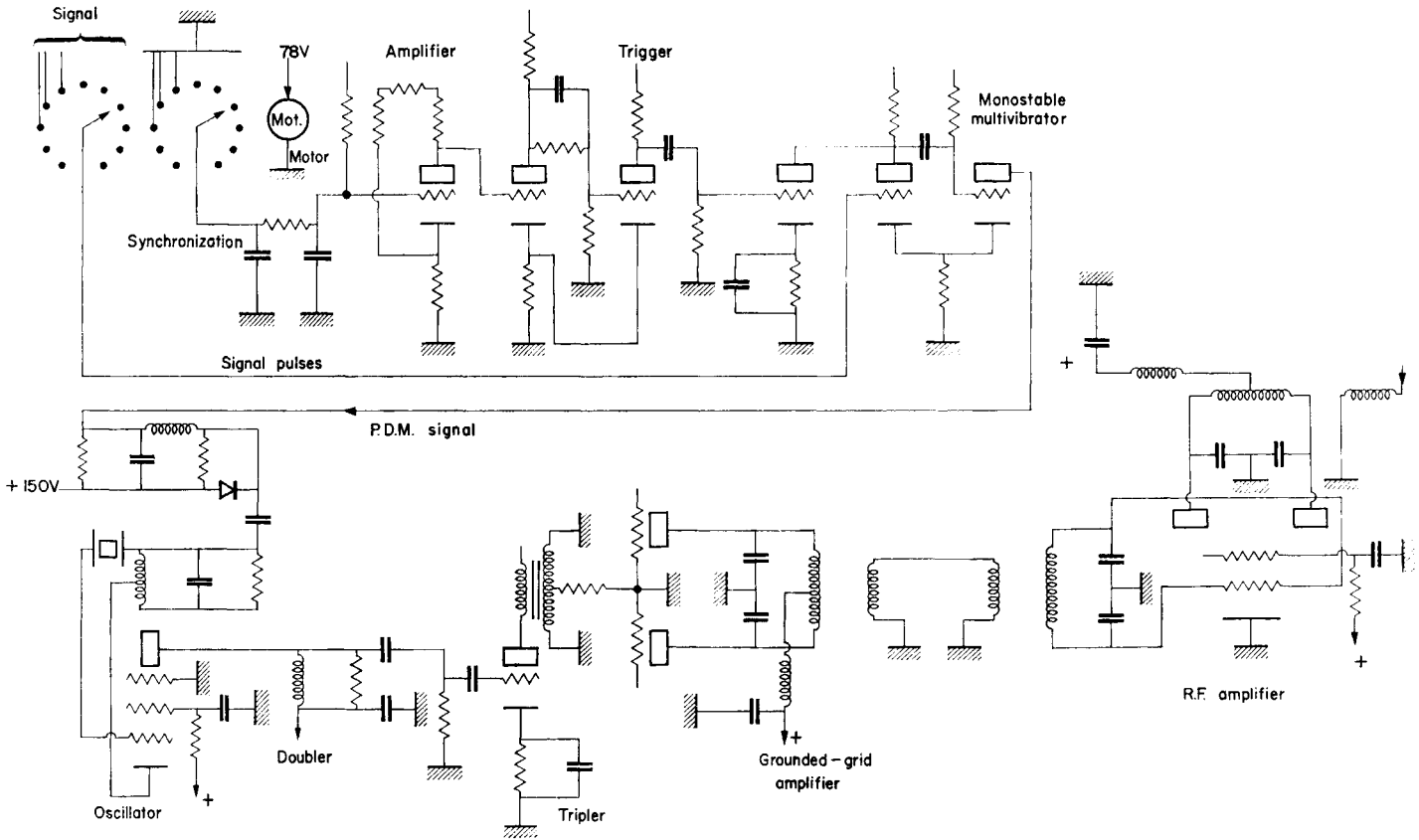


FIG. VIII.29. A PDM–FM transmitter[72].

The synchronization pulses have a repetition frequency of 750 c/s and are distinguished from the rest by a momentary increase in the amplitude of the carrier wave.

The multiplexer is the one that was described in VIII.4.3. The pulses which leave the multiplexer are applied to the input of the coder (see Fig. VIII.31). The pulses have been lengthened beforehand, to 41·6 μs (=1/24,000) as this corresponds to one whole sample period.

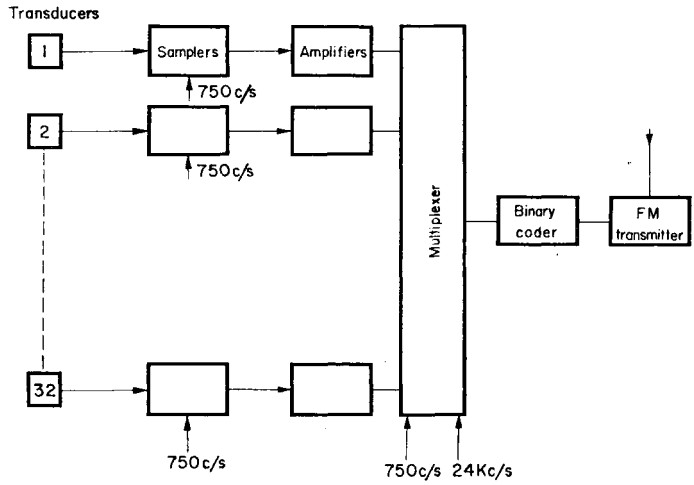


FIG. VIII.30. Block diagram of complete AN–AKT/14 transmitter.

We obtain at the output of a series of words containing eight active digits together with a "0". A digit "1" appears as an amplitude of +5 V, and a digit "0" as −5 V.

A sample of information is successively compared with the weighted binary currents provided by the coder.

If the input signal is smaller than half the total scale, the coder indicates "0" and a second comparison, this time with one quarter of the total scale, is then made. If the input signal is now higher than this level, the coder indicates the digit "1". The process continues through the 8 comparison levels (see Fig. VIII.32).

The currents through the weighted current gates take up their true values when the point *A* reaches earth potential.

The signal current and the weighted current at the output of the decoder are added together at the amplifier input by the resistor *R*.

The error amplifier detects a signal which may be positive, negative or zero; if it is negative or zero, the "AND" gate is not effected and the flip-flop does not give a digit. When the "AND" gate receives a positive signal, however, the flip-flop is triggered and gives the digit "1".

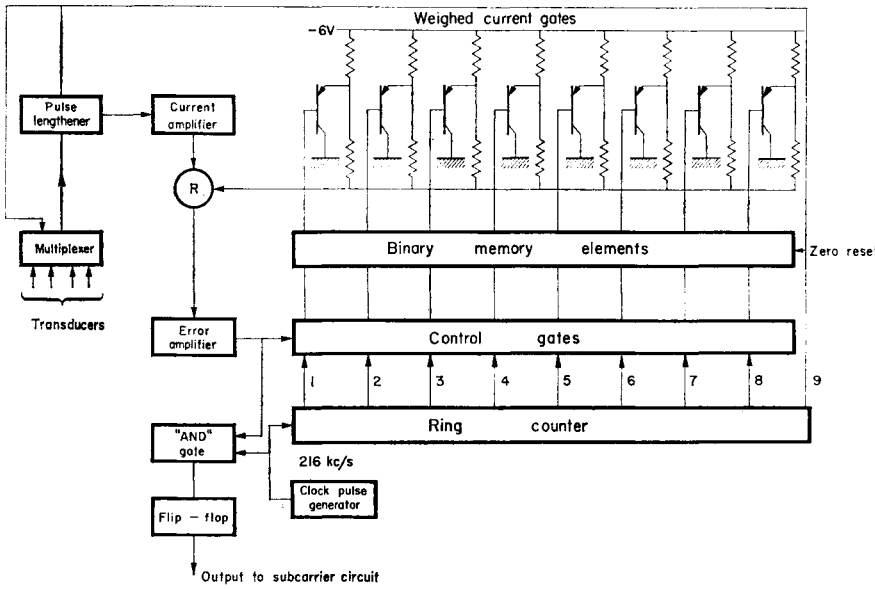


FIG. VIII.31. The AN–AKT/14 coder.

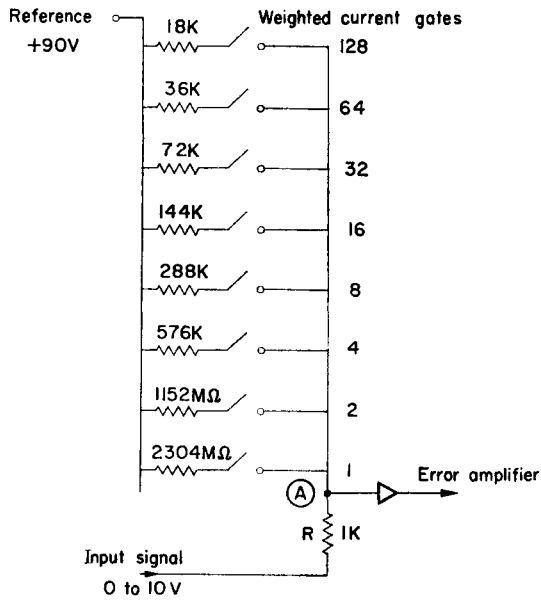


FIG. VIII.32. Comparison circuit.

With each timing pulse, the ring counter makes the control gates advance by one digit, and a memory circuit retains the preceding digits throughout the sequence. The signal current and the decoder current are thus compared sequentially. At the end of each sequence, a counter pulse resets the circuits to "0" and the whole procedure begins again for the following sample.

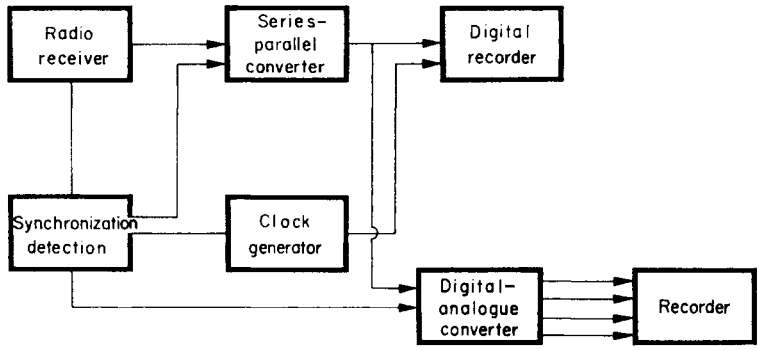


FIG. VIII.33. Block diagram of complete UKR 7 receiver.

The UKR–7 receiver is shown in Fig. VIII.33. A 216 kc/s oscillator (216 kc/s = 9 × 24 kc/s) provides 750 c/s after division; it is controlled by the frequency difference between the oscillator and the synchronization detector. When a frequency error is detected, a voltage is produced which is applied to a reactance valve in the oscillator circuit.

The various multiplex channels are detected in series, and a series-parallel converter redistributes the telemetry channels into parallel form. The signals from each channel appear in digital form, and a digital-analogue converter reconstitutes the data in its original analogue form.

VIII.5.3. *AJAX, an Example of an FM–PM Telemetry Device*

The AJAX telemetry device consists of an FM–PM multiplex unit. The data supplied by the detectors are applied to transducers which provide frequency-modulated sub-carriers. This mixture of sub-carriers is applied to a carrier modulator. Each sub-carrier can be modulated directly by the signal from a detector, or by the output signal from a commutator; the latter enables a greater number of channels to be explored (PAM–FM–PM transmission).

With AJAX, sixteen sub-carriers can be transmitted simultaneously, which correspond to the fifteen channels of a standard IRIG; the sixteenth channel which has a nominal frequency of 125 kc/s is peculiar to the AJAX system.

The modulation factors of the sub-carriers are usually 7·5 per cent. The carrier frequency lies in the 250 Mc/s range. A transmitter of the AJAX variety is shown schematically in Fig. VIII.34.

The instruments on board which have been specially adapted for missiles are miniaturized; they comprise the following units:

detectors;
oscillators coupled to the detectors;
potential transducers;
amplifiers which provide insulation and adaptation;
a mixer amplifier which provides the multiplex signal;
a HF emitter which is phase-modulated by the multiplex signal.

The receiving equipment consists of:

aerials followed by pre-amplifiers. Multicouplers can be used if necessary, to allow several VHF receivers to be coupled to a single aerial;
a demodulation unit, which reconverts the signals into LF at the output;
a checking and simulation unit, which controls the whole system.

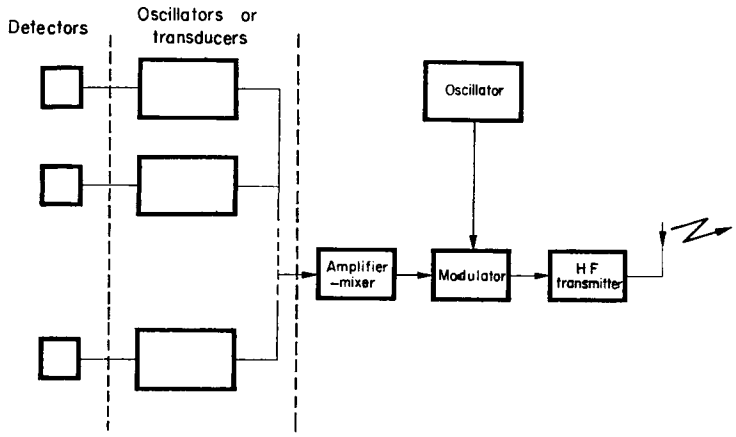


FIG. VIII.34. The transmission system of the AJAX telemetry device.

The receiver and the demodulation unit consist of the VHF receiver together with the necessary components for separating and demodulating the twelve telemetry channels.

It is possible to record the multiplex signal on magnetic tape. The whole system is illustrated schematically in Fig. VIII.35.

The field which is received at the aerial is separated into two vectors, which correspond to fields circularly-polarized clockwise and anti-clockwise. The two HF channels are amplified separately and combined. The ensuing signal

is applied to a demodulator which reproduces the signals in the modulated sub-carriers. These output signals can be used directly using analogue techniques; alternatively, they can be recorded, or once the transmitted signal has been commutated, the output signal can be extracted to furnish the original message. In this final situation, the decommutation unit detects the PAM,

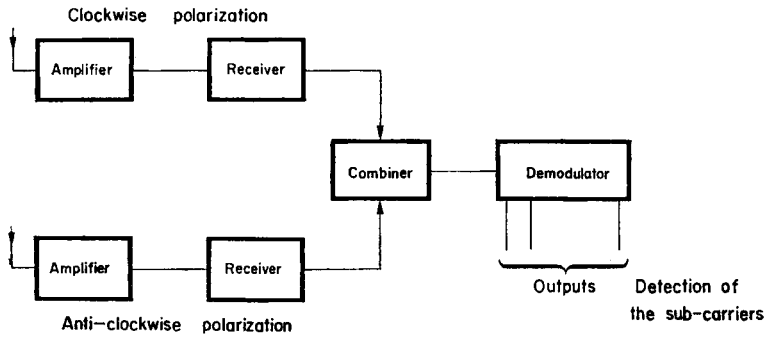


FIG. VIII.35. The layout of an AJAX receiver.

PDM or PCM synchronization signals and performs the decommutation, that is, performs a series-parallel transformation of the selected measurement channels.

VIII.6. RECOVERY OF THE DATA

After detecting and filtering the subcarrier waves, the LF channels are extracted either by further detection or by discrimination; in the case when they are simply transposed, these channels can then be reconstituted by filtering.

When the LF channels have been distributed by sampling, the decommutator at the receiver provides the data from the channels in parallel form.

Once the telemetry channels have been detected, the data contained in each is recorded.

VIII.6.1. *Recording Devices*

As soon as the data have been reconstituted in the form of a voltage, they is fed into a recording device. This may use a pen and ink, in which case the recording is made on paper, or it may alternatively be made upon photographic paper (see Fig. VIII.36). This type of recorder employs a galvanometer movement incorporating a mirror. A time-counter allows the time scale to be calibrated as the film unwinds.

As an example, consider a typical SFIM recorder. The photographic paper is wound onto a reel by a motor which is speed-controlled by a thyratron. A

wide range of speeds, between 3 mm/sec and 3 m/sec is available. The device is so arranged that the recorder can be controlled from a distance. The variation between the speeds of rotation when fully loaded and empty is 3 per cent. The speed is not affected by a variation of ± 1 per cent in the mains voltage.

A maximum of 25 channels can be recorded. The fineness of the recorded spot is a function of its brightness and the writing speed. For a 15,000 c/s recording, with a winding speed of 2 m/sec, the line is $\frac{4}{10}$ mm wide.

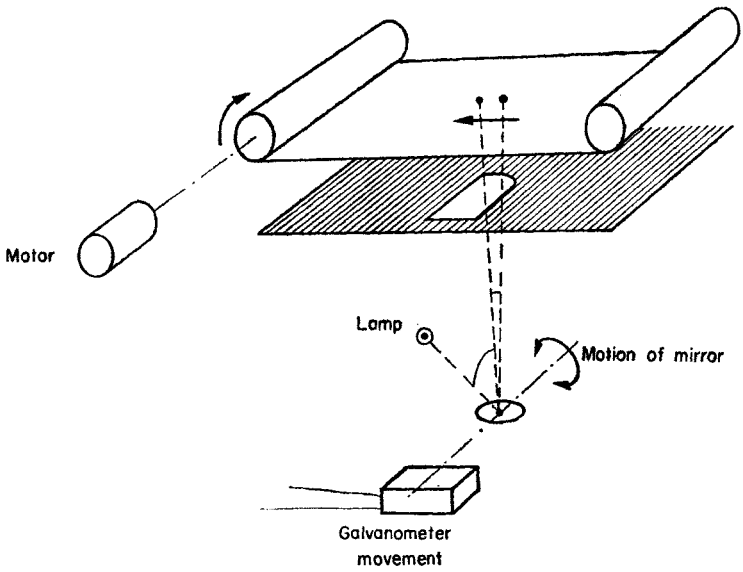


FIG. VIII.36. Photographic recording.

VIII.6.2. *Magnetic Recording*

This procedure is being used more and more. It is possible to record frequencies higher than 100 kc/s with little distortion under ordinary conditions. In special cases, even higher frequencies can be recorded on magnetic tape. This system is particularly well adapted for use with digital data; it is only necessary to record the subcarrier frequencies directly, and the information in the various channels is reconstructed automatically by means of a calculating machine. By doing this, problems of interpretation posed by recordings made on paper (numerous lines recorded) are avoided.

One drawback of magnetic recorders is flutter; this phenomenon can introduce errors of some 6 or 7 per cent into the measurements.

The signal to be recorded may be direct or alternating, and its frequency

low or high, and at a low level; it may also consist of transient phenomena. Magnetic tape has the attraction of possessing flexibility in recording.

There are three main recording methods: direct, with frequency modulation and with carrier erasure.

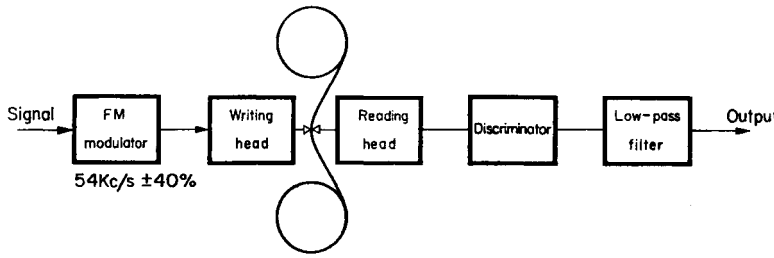


FIG. VIII.37. FM recording device, using magnetic tape.

In direct recording, the signal is amplified to the required level, and mixed with a relatively high frequency; the resulting signal is applied to the recording head. In this way, frequencies between 50 c/s and 250 kc/s can be recorded. The signal-to-noise ratio is 35 db for 1 per cent distortion of the third harmonic; for 3 per cent distortion, the signal-to-noise ratio becomes 40 db.

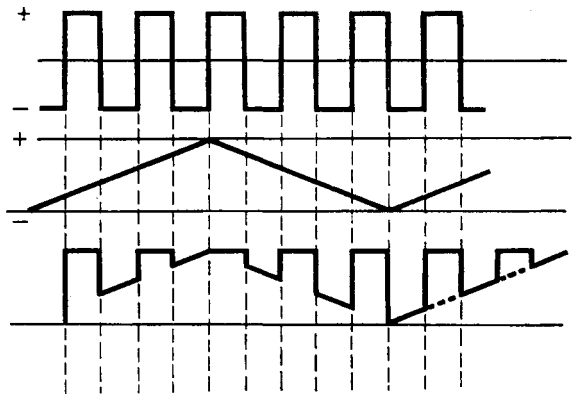


FIG. VIII.38. The principle of recording by carrier wave erasure.

The direct recording process is limited by the amplitude variation caused primarily by non-uniformity of the magnetic material. It does, however, enable a relatively wide frequency band to be covered, and it is now possible to record frequencies of the order of a megacycle in special cases. With frequency modulation recording, the problem of amplitude variation is avoided and the lower limit of the recording is extended to d.c. A fixed carrier wave

which is modulated in frequency by the signal is used. A diagram of the device is given in Fig. VIII.37.

This arrangement is particularly suitable for recording FM–FM multiplex signals; it has the advantage of responding to very low frequencies (a steady quantity can be recorded), and time expansion of the signal can be obtained which is favourable to the signal-to-noise ratio.

Recording with carrier-wave erasure is used in miniature recorders, where the electronic equipment has to be kept down to a minimum. The method results in a form of AM signal. A carrier wave is recorded beforehand onto the tape, and the useful signal erases this carrier by an amount proportional to its amplitude (see Fig. VIII.38).

The data is recovered by rectifying and filtering the output voltage.

In this way the input signal can be recorded at one speed and reconstructed at another (for example recopied on a pen-recorder). Further the timing signal is obtained at the same time.

Six speeds are usually available: 4·76, 9·52, 19·05, 38·1, 76·2 and 152 cm/s; the choice depends upon the frequency of the signal to be recorded.

The tape width is 25·4 mm. About 14 tracks can be used with this width.

Apart from the standard applications of magnetic recorders, miniature recorders should be mentioned, such as the MTR 1200 made by the *Leach* company-which can store data during flight aboard a missile, and this data can subsequently be reconstructed by the remote-control of a telemetry device.

CHAPTER IX

CONCLUSION

REMOTE-CONTROL and telemetry have many points in common, as we have seen in the first six chapters, and devices particularly associated with one and the other have been studied separately in the last two chapters.

Every effort has been made not to go beyond the limits laid down and defined by the title of the book. We have by no means exhausted the field of telemetry and remote-control, either as regards missiles or in the more general framework of interplanetary communications. These systems may also find applications in automation and cybernetics, and the use of radar techniques in connection with measurement opens up a very large field.

It is, for example, possible to measure an altitude by means of a radar device known as a radio-altimeter or altimeter probe. Data concerning the altitude can be retransmitted by telemetry, but with some types of equipment the altitude can be received directly on the ground, without the use of a supplementary telemetry device.

These techniques have been employed ever since the last war, both with pulse and frequency modulation. The altimeter probe with frequency modulation seems to be more flexible, especially where flight at low altitude is concerned. It is possible to link the controls of the missile to the altitude measurement provided by the probe in such a way that the missile can be flown at a constant altitude. This is equivalent to using a telemetry device for guiding the missile; it is an interesting example of a combination of remote-control and telemetry.

This automatic control procedure has been further developed by different techniques; it is related to the various control systems.

Another application of telemetry by means of radar concerns measurement of the distance between a missile and its target, either by an active or passive method. In the latter, the measurement is carried out only at the target, and there is no responder aboard the missile. Such measurements as these call for Dopller radar, or frequency-modulation radar. The data about the distance still to be traversed are retransmitted to the ground either by telemetry or directly.

This distance measurement may also control the detonation of the missile, resulting in a device known as a minimum distance firing indicator.

There is an ever-growing tendency to group telemetry, remote-control and position-finding into one single unit.

The identification transmitter LIRNA of Nord-Aviation, for example, not only allows us to track a target aircraft from a fixed station, but also to obtain the altitude and the speed of the missile through two telemetry channels in the responder which is on board.

The TELEBIT system, which was used by the United States in the interplanetary rocket *Explorer VI* contains two sub-units, for telemetry and remote-control, which are connected to the same aerial through a duplexer.

It should be mentioned also that television pictures can be transmitted by telemetry.

Reconnaissance missiles form a large topic on their own; instantaneous exposures, made over some region, can be stored in memories and retransmitted at will by telemetry; the timing of the operations is directed either by a programme or by remote-control.

We can envisage how remote-control and telemetry might find applications in the wider sphere of automatic control and measurement.

Today, many industries are turning towards automation, not only for automatic measurement at check points during manufacture but also for fully-automatic production lines. Telemetry and remote-control play a predominant role in this.

While undoubtedly automation in factories has so far only affected a relatively limited area, and radio links are of little importance, it is by no means unthinkable that, in the more or less remote future, automation will develop in factories in such a way that for each type of industry, a standard installation will be produced.

It is clear that coordination between similar factories, or factories dependent upon one another, must exist: this connection represents the existence of a programme, of a general plan. From this point of view, a radio link has undisputed advantages.

To come closer to present day reality, it is easy to envisage the value of coordination by means of radio remote-control and telemetry where an automatic factory is concerned. With their aid, we can deal more flexibly with the installation, and the machinery can be distributed according to the amount of work it has to do.

Remote-control and telemetry equipment has, therefore, a considerable role to play in industry.

The production of missiles and spacecraft is no longer speculative, but forms, rather, an *avant-garde* industry.

Automation, too, is going to revolutionize present day techniques, and I believe that telemetry and remote-control must fill an important place in that they form particular elements in the overall system. There is, moreover, a resemblance between the guidance sequence of a missile and the transfer sequence in automation; the nature of the programme alone differs; information theory, optimization by filtering, the stability of the feedback loops and the energy sources are all basic features.

Other examples of future applications of telemetry and remote-control devices are the automation of radio multiplex systems or telephone connections, distant signalling in large-scale production plants, the distribution and employment of energy, and the remcte-control or self-control of atomic regulators. Further perspectives are opened up by micro-miniaturization of the equipment, in medical research in particular; telemetry transmitters in the form of pills already exist in the United States.

APPENDIX

STANDARDS FOR TELEMETRY

(Document I.R.I.G. No. 103-56, Oct. 1956)

The committee of the "Department of Defense Board on Guided Missile Research and Development", has standardized telemetry frequencies in the United States. In this appendix we reproduce the frequency values which have been adopted in the United States, in the hope that this standardization may become more general.

RF frequencies. Two frequency bands have been allocated to telemetry:

$$216 < f < 235 \text{ Mc/s}$$

and

$$2200 < f < 2300 \text{ Mc/s}.$$

The use of FM–FM and FM–PM telemetry without channel commutation. A maximum of 18 subcarriers are used, the characteristics of which are collected together in the following table:

TABLE A.1

Band number	Central frequency (c/s)	Lower limit	Upper limit	Maximum deviation (%)	Frequency response (c/s)
1	400	370	430	±7·5	6
2	560	518	602	—	8·4
3	730	675	785	—	11
4	960	888	1032	—	14
5	1300	1202	1398	—	20
6	1700	1572	1828	—	25
7	2300	2127	2473	—	35
8	3000	2775	3225	—	45
9	3900	3607	4193	—	59
10	5400	4995	5805	—	81
11	7350	6799	7901	—	110
12	10,500	9712	11,288	—	160
13	14,500	13,412	15,588	—	220
14	22,000	20,350	23,650	—	330
15	30,000	27,750	32,250	—	450
16	40,000	37,000	43,000	—	600
17	52,500	48,560	56,440	—	790

TABLE A.1 (contd.)

Band number	Central frequency (c/s)	Lower limit	Upper limit	Maximum deviation (%)	Frequency response (c/s)
18	70,000	64,750	75,250	—	1050
A	22,000	18,700	25,800	±15	660
B	30,000	25,500	34,500	—	900
C	40,000	34,000	46,000	—	1200
D	52,500	44,620	60,380	—	1600
E	70,000	59,500	80,500	—	2100

For a maximum of only 14 channels, channels 2 to 15 are used.

If we use the band	we omit the bands
A	15 and *B*
B	14, 16, *A* and *C*
C	15, 17, *B* and *D*
D	16, 18, *C* and *E*
E	17 and *D*.

The frequencies must be accurate to within an error of ½ per cent.

It is interesting to notice that the document recommends de-emphasis greater than 9 db per octave.

Commutation of the channels. Commutation may be employed on one or several of the subcarrier waves. The commutation speed on each subcarrier is limited to a certain value (see Table A.2).

Conservative value. This is the value of the sampling time, chosen to be of such a length that the reading error during the second half of the sampling – due to transient effects when commutation occurs – does not exceed ½ per cent of the quiescent value. These values are applicable when the recording equipment is limited by a low-pass filter, the cut-off frequency of which is equal to the frequency response specified in Table A.I.

The "minimum value" of the sampling time is one quarter of the "conservative" value. These values should be applied only for fast sampling speeds and only when the low-pass filters (for deviation ratios of 5) have been replaced by elements which have cut-off frequencies equal to four times the recommended frequency response.

Decommutation of the channels. The total number of samples per turn (the number of contacts on the mechanical commutator) is given in Table A.3.

– The commutation diagram is given opposite in Fig. A.1.

– The synchronization pulse is equal to twice the period of contac plus a cutoff period on each sequence.

– The speed of the commutator (or of the sequence) should not vary by more than +5 to −15 per cent of the nominal value given in Table A.3.

TABLE A.2

Band number	Central frepuency (c/s)	Sampling time (msec)		Commutation speed (samples/sec)	
		Conservative value	Minimum value	Conservative value	Maximum value
1	400	670·0	170	1·5	6
2	560	480	120	2·1	8·4
3	730	370	91	2·7	11
4	960	280	70	3·6	14
5	1300	210	51	4·9	20
6	1700	160	39	6·4	25
7	2300	120	29	8·6	35
8	3000	89	22	11	45
9	3900	68	17	15	59
10	5400	49	12	20	81
11	7350	36	9·1	28	110
12	10,500	25	6·4	39	160
13	14,500	18	4·6	55	220
14	22,000	12	3	83	330
15	30,000	8·9	2·2	110	450
16	40,000	6·7	1·7	150	600
17	52,500	5·1	1·3	200	790
18	70,000	3·8	0·95	260	1050
A	22,000	6·1	1·5	170	660
B	30,000	4·4	1·1	230	900
C	40,000	3·3	0·83	300	1200
D	52,500	2·5	0·63	390	1600
E	70,000	1·9	0·48	530	2100

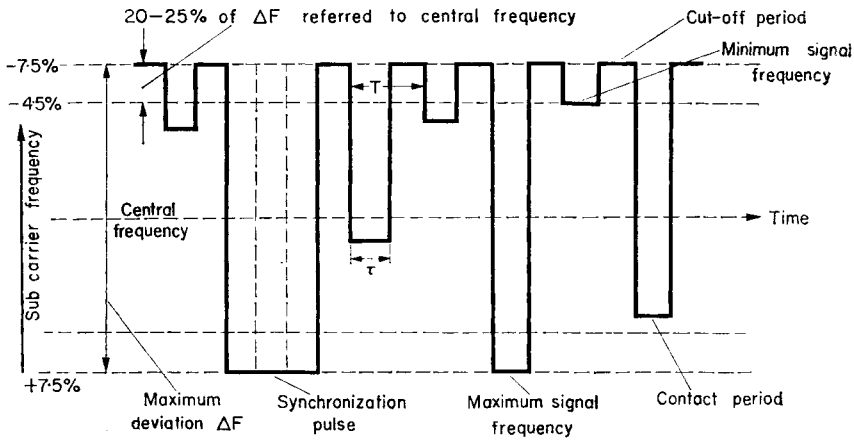


FIG. A.1

TABLE A.3

Number of samples per sequence	Sequential speed (number of sequences per second)	Commutation speed (number of samples per second)	Minimum bandwidths for the recommended subcarrier waves
18	5	90	14,500 c/s
18	10	180	22,000 ± 15% or 30,000 ± 7%
18	25	450	30,000 ± 15% or 70,000 ± 7%
30	2·5	75	10,500
30	5	150	22,000 ± 7·5%
30	10	300	22,000 ± 15% or 40,000 ± 7%
30	20	600	40,000 ± 15%
30	30	900	70,000 ± 15%

- If coincidence gating is employed on the aircraft, the length of each information pulse lies between 47 and 53 per cent of the period of the channel, and should be stable to within $\pm\frac{1}{2}$ per cent of the nominal period of the channel. It should be noted that the stability of the duration of the information pulse will affect the accuracy.
- When there is no coincidence on the aircraft, it must be produced on the ground, and the ratio of the length of the information pulse to the channel period must lie between 60 and 75 per cent.

Radio frequency. From 216 Mc/s to 235 Mc/s
- band occupied by a telemetry signal: 0·5 Mc/s
- RF stability: ±0·01 per cent
- frequency deviation: ± 75 kc/s $< \Delta f <$ 125 kc/s
- power: 100 W maximum, but this depends upon the threshold distance.

The PDM–FM telemetry standard. This arrangement is used for a great many channels, with a frequency response which is, however, lower than FM–FM channels.

PDM specification
- number of samples per sequence : 30 45 60 90
- sequential speed (seq/sec) : 30 20 15 10
- commutation speed (samples/sec) : 900 900 900 900
- the relation between the amplitude and the pulse duration must be linear.
- minimum pulse duration (zero level information) : 90 ± 30 μsec
- maximum pulse duration (maximum level information) : 700 ± 50 μsec

– rise-time and fall-time : 10 to 20 μsec

(measured between 10 and 90 per cent of the levels) constant to within ±3 μsec for a given transmitter

– commutator speed : +5 to −15 per cent of the nominal value.

Figure A.1 shows the diagram for PDM commutation.

Specification for the Carrier Wave

– RF frequency: $216 < f < 235$ Mc/s
– bandwidth occupied: 0·2 Mc/s
– stability: ±0·01 per cent
– modulation: FM(PDM–FM)
– frequency deviation: ± 25 kc/s $< \Delta f < \pm 45$ kc/s
– maximum power: 100 W
– power of the harmonics: to −60 db of the fundamental.

Standards for PCM Telemetry (IRIG No. 102–59)

Definitions

– *synchronization*: the beginning of a sequence is indicated by a chosen code word
– *subcommutation*: the process whereby two or more multiplex systems are placed in cascade
– *supercommutation*: the use of several channels for a measurement in a system with time division
– *sequence*: a recurrent number of words, comprising a single synchronization signal
– *speed of the sequence*: the frequency derived from the period of the sequence
– *parity check*: a code automatically checks the total number of "1" (or of "0") which is always odd or even
– *word*: ordered collection of bits
– *word speed*: the frequency derived from the period between the beginning of the transmission of a word and the beginning of the transmission of the following word.

The transmission is carried out in the binary system.

Information Rate as a Function of the RF Bandwidth at 3 db

bits/sec	c/s
50,000 ± 10%	75,000
100,000 ± 10%	150,000
200,000 ± 10%	300,000
400,000 ± 10%	600,000
800,000 ± 10%	1,200,000

The RF bandwidth at the receiver equals 1·5 times the information rate.

Word Structure

- a word will consist of at most 15 bits
- the number of bits/word will be constant during a continuous transmission
- the bits will be transmitted in order of decreasing importance
- a bit for checking the parity will be included in each word, and will be such that the sum of all the "1" in a word is odd
- *words per sequence*: the number of words per sequence will not exceed 130
- words per second (word speed): the total number of words per second, including all the words used for the synchronization, will not exceed 60,000.

Modulation of the RF Carrier Wave—FM or PM

(a)—*recommended method*: a digit "1" is represented by a deviation of the RF carrier wave in the direction of higher frequecnies, and a "0" by a deviation of the RF carrier wave in the direction of the lower frequencies

(b)—*a possible method*: a digit "1" is represented by the presence of a deviation of the RF carrier wave, and a digit "0" by its absence.

BIBLIOGRAPHY

1. ATWOOD, S.S. and BURROWS, C.R. *Radio Wave Propagation*, Academic Press, N.Y. 1949.
2. ARNIER, G. L'écoute des satellites artificiels au centre RTF de Limour, *Onde Elect.*, July–Aug. 1959.
3. BAIN, W.C. and SHEARMAN, E.D.R. Observations on the USSR earth satellites and the study of radio wave propagation. *PIEE*—Part B, May 1959.
4. BAKER, R.C. A circularly polarized feed for an automatic tracking telemetry antenna, *IRE Trans. on Space Electronics and Telemetry*, Sept. 1959.
5. BEENKEN, H.G. and DUNN, F.L. Short distance radio telemetry of physiological information, *IRE Trans. on Space Electronics and Telemetry*, June 1959.
6. BERARD, A. Les appareils de mesure à seuil, *Electronique*, May 1959.
7. BENECKE, TH. and QWICK, A.W. *History of German guided missiles development*, Brunswick, Germany 1957.
8. BIGELOW, G.F. and TAYLOR, L.S. Channel selection for multicarrier telemetry, *I.R.E. Conv. Rec.*, Part 5, U.S.A. 1958.
9. BISHOP, R.P. and MARQUAND, E.F. The development of a high speed electronic multiplexer and coder for use with a PCM telemeter, *IRE Conv. Rec.*, Part 1, 1956.
10. BORG, H. Théorie des circuits impulsionnels, *Revue d'Optique*, 1953.
11. BRAUN, E.L. and GIANOPLUS, A.S. A digital computer system for terminal area air traffic control, *IRE Trans. on Space Electronics and Telemetry*, June 1959.
12. BRUINSMA, Télécommande par radio, *Bibliothèque technique Philips*, Dunod, 1954.
13. BRUNO, J. Télémesure SECRE quantifiée et codée, *Electronique*, June 1956.
14. BURBERRY, R.A. Aerial systems for aircraft, *J. Royal Aeronaut. Soc.*, Feb. 1956.
15. BURGESS, E. *Guided Weapons*, Chapman & Hall, London 1957.
16. BURROWS, C.R. see reference No. 1.
17. CABESSA, R. Liaison transhorizon—Utilisation et performances, *Onde Elect.*, Jan. 1960.
18. CHATELAIN, G. Les antennes dans l'espace, *Onde Elect.*, Oct. 1959.
19. CHATTERTON, J.B. The uncertainty of measurement systems, *Inst. Radio Engrs. Trans.* I, March 1958.
20. CLARK, R.C. Diagrammation methods of code construction, *Communic. and Electron.*, Jan. 1959.
21. CLEMENT, A. and BRUNO, J. Les télémesures cycliques quantifiées et codées SECRE, *Onde Elect.*, March 1957.
22. COGAN, S. and HODDER, W.K. A high speed airborne digital data acquisition system, *IRE Trans. on Space Electronics and Telemetry*, Sept. 1959.
23. COUILLARD, L. and GIRAUD, P. Sondeur radio pour la measure des hauteurs des aéronefs au-dessus du sol, *An. Rad.*, April 1947.
24. DEMAN, P. Spectre instantané et analyse du signal simultanément en fréquence et en temps, *An. Téléc.*, Jan./Feb. 1959.
25. DESPORTE, Les télémesures en Aéronautique, Rapport d'étude No. 207–CEV–1956.
26. DEZOTEUX, M. Facteur de bruit des transistrons à jonctions, *Electronique*, 15 March 1958.
27. DJORUP, R.S. Comments relative to the application of PCM to aircraft flight testing, *IRE Trans. on Space Electronics and Telemetry*, Sept. 1959.
28. DUNN, F.L. see reference No. 5.

29. EBERHARDT–RECHTIN Communication techniques for space exploration, *IRE Trans. on Space Electronics and Telemetry*, Sept. 1959.
30. ECKHARDT, D. Equipement de télémesure pour fusée ou ballon-sonde, *Onde Elect.*, July/August 1959.
31. EDDINS, W.T. Circuits d'échantillonnage pour corrélation temps d'information à tension analogique, *IRE Nat. Conv. Rec.*, Pt. 5, 1958.
32. FAGOT, J. Transmission par faisceaux hertziens, *Onde Elect.*, Jan. 1956.
33. FAGOT, J. Causes diverses de diaphonie dans les systèmes multiplex à impulsions, *An. Rad.*, Oct. 1953.
34. FEUNG, T.N. Four ways to specify non-linearity of linear transducer-control, *Engng. U.S.A.*, May 1959.
35. GAUDILLÉRE, P.C. Le guidage des engins, *Fusée Rech. aéron.*, April 1958.
36. GIANOPLUS, A.S. see reference No. 11.
37. GILCHRIST, C.E. Application of the phase locked loop to telemetry as a discriminator or tracking filter, *Inst. Rad. Engrs. Trans. TRC.*, June 1958.
38. GIRAUD, P. see reference No. 24.
39. GOUDET, G. Les télécommunications par ondes centimétriques, *Rev. d'Optique*, 1950.
40. GRANDSARD, P. Augmentation de la surface equivalente des engins-cibles, Internal note, Nord-Aviation No. 2571.
41. GRAY, A.R. Reliability of missile guidance systems, *I.R.E. Conv. Rec.*, Part 6, 1958.
42. GREER, J.R. A sensitive defocusing photoelectric pressure transducer, *Electron. Engng.*, July 1958.
43. GRISETTI, R.S. and MULLEN, E.B. Base-line guidance systems, *IRE Trans.*, Dec. 1958.
44. GUTTON, H. Pilotage d'un astronef par des moyens radioélectriques, *8ème Congrès International d'Astronautique de Barcelonne*, 1957.
45. HODDER, W.K. see reference No. 23.
46. JURGEN, R.K. How transducers measure and control, *Electronics*, 4 July 1957.
47. KELLEY, G.J. Choosing the optimum type of modulation—A comparison of several communication systems, *Inst. Rad. Engrs. Trans. C.S.*, June 1958.
48. KENNEDY, W.F. see reference No. 81.
49. KLASS, P.J. Airborne spiral antennas minimize drag, *Aviation Week*, 14 July 1958.
50. KRAMER, A.S. Electronic commutator in multiplex telemetring, *Electronics*, Sept. 1959.
51. KRAUS, J.D. Les antennes hélicoïdales et leurs applications aux larges bandes, *PIRE*, Oct. 1948.
52. LEGETTE, M.A. Strain-gauge principles, *Instrum. Automation*, March 1958.
53. LIBOIS, L.J. *Faisceaux hertziens et systèmes de modulation*, Chiron 1958.
54. LOCKE, F.A. A digital data transmission system using phase modulation and correlation detection, Hughes Aircraft Company, U.S.A.
55. LOSEE, F.A. *Guidance*, Van Nostrand, U.S.A., 1956.
56. LOYEN, M. Procédés de modulation utilisés dans les liaisons entre la terre et les engins spatiaux, *Journées d'Information astronautique*, Vol. II, May–June 1960.
57. LUCKE, W.H. A magnetic amplifier commutating and pulse-width encoding circuit, *Communic. and Electron.*, Jan. 1959.
59. MARCUS, J. *La modulation de fréquence: théorie et applications industrielles*, Eyrolles, 1960.
60. MARCUS, J. Antennes spirales, Internal note 3752, ELAE, Nord-Aviation.
61. MARTIN, A.V.J. Note sur la réduction de la quantité d'information contenue dans une image de télévision, *Onde Elect.*, July 1959.
62. MATTHEWS, W. Telemetring in earth satellites, *Electr. Engng.*, Nov. 1957.
63. MERTZ,P. Information theory impact on modern communication, *Communic. Electron.*, Sept. 1957.
64. NICHOLS, M.H. and RAUCH, L.L. *Radio Telemetry*, 2nd ed., Wiley, New York, 1956.
65. OLIVER, B.M. Efficient coding, *Bell Syst. Techn. J.*, July 1952.

66. ORTUSI, J. La propagation des ondes métriques et centimétriques—Détermination de l'équivalent de transmission, *An. Rad.*, July 1964.
67. ORTUSI, J. L'aspect physique de la théorie de l'information—application au choix des liaisons multiplex et en télévision, *An. Rad.*, Oct. 1957.
68. PETTINGALL, C.E., SINK, R.L. and POSTHILL, B.N. An automatic data handling system of high speed and high capacity, *IRE Trans. on Space Electronics and Telemetry*, March 1959.
69. POINCELOT, P. Influence de l'absorption sur le coéfficient de réflexion de l'ionosphère, *An. Téléc.*, March–April 1959.
70. POSTHILL, B.N. see reference No. 68.
71. MCRAE, D.D. Consideration of RF parameters for PCM telemetry systems, *IRE Trans. on Space Electronics and Telemetry*, June 1959.
72. RASKHODOFF, N. Telemetry transmitter for Vanguard rocket, *Electronics*, 4 July 1958.
73. RAUCH, L.L. see reference No. 64.
74. RICHTER, H. Système de repérage radioélectrique microlock—Jet Propulsion Labor. CIT.
75. ROLOFF, H.A. Telemetring receiving system at the Air Force Missile Test-Center, *Inst. Rad. Engrs. Trans.*, Dec. 1957.
76. ROUBINE, E. Lignes et Antennes, *Revue d'Opt.* 1954.
77. ROWE, W.D. A transistorized digital-to-analog converter, *IRE Trans. on Instrumentation*, March 1958.
78. RUSCUS, J.C. and WRIGHT, J.L. A superfast carrier quieted telemetry system for missile impact instrumentation, *IRE Trans. on Space Electronics and Telemetry*, June 1959.
79. SCHARLA–NIELSEN, H. Technique for interplanetary telemetring, *Proc. Nation. Conf. Aeron. Electron.*, May 1958.
80. SCHARLA–NIELSEN, H. Spaceship telemetry, *Inst. Rad. Engrs. Trans.*, June 1958.
81. SCHENCK, J. and KENNEDY, W.F. Analysis of multiplex error in FM/FM and PAM/FM/FM telemetry, *IRE Trans. on Space and Telemetry*, Sept. 1959.
82. SHANAHAN, E.F. Problems in aircraft telemetring, *Inst. Rad. Engrs. Trans.*, Dec. 1957.
83. SHANNON, C.E. Certain results in coding theory for noisy channels, *Inform. and Control*, Sept. 1957.
84. SHAW, G.S. The AN/AKT 14 Telemetry system, *IRE Trans. Telemetry and Remote Control*, March 1956.
85. SHEA, R.F. *Transistor Circuit Engineering*, Wiley, New York, 1957.
86. SHIMMINS, A.J. Radio Telemetry—Part I—Systems, *Proc. Inst. Rad. Engrs. Austral.*, Dec. 1958.
87. SINK, R.L. see reference No. 68.
88. SMITH, C.H. Future requirements for range telemetry, *Proc. Nation. Conf. Aeron. Electron.*, May 1958.
89. SOUBIES–CAMY, H. Le traitement des informations codées, *Automatisme*, March–April–May 1959 and Nov. 1959.
90. SUSSKIND, A.K. *Analog-Digital Conversion Techniques*, Chapman & Hall, N.Y., 1956.
91. SWAIN, W.H. Transistorized subcarrier oscillator principles, *1959 Nat. Symp. on Space Electronics and Telemetry*, Sept. 1959.
92. SWERLING, P. Space communication, *IRE Trans. on Military Electronics*, Dec. 1958.
93. TAYLOR, L.S. see reference No. 8.
94. TELTSCHER, E. and WEISSMAN, L. A pulse position telemetry system, *IRE Conv. Rec.*, Part 5, 1958.
95. TEPPER, M. *Fundamentals of Radio Telemetry*—Fasc. No. 225 of the collection: Rider Publications—Rider N.Y. 1959.
96. TERMAN, F.E. *Radio Engineers Handbook*, McGraw-Hill, 1956.
97. THOUREL, L. *Antennes aeroportées*—Cours professé à l'ENSA, 1960.
98. THOUREL, L. *Les antennes*, Dunod, 1956.

99. TURNER, F.M. The impact of the spiral antenna and its derivatives on the design of antennas and weapons systems, *Aeron. Electronics National Conf. Proc.*, May 1958.
100. VASSEUR, J.P. Impulsion de Gauss, *An. Rad.*, Oct. 1953.
101. VASSEY, E. L'émission radioélectrique des satellites artificiels, *Onde Elect.*, July–Aug. 1959.
102. VOELCKER, H.B. *Phase Shift Keying in Fading Channels*, Inst. of Electr. Eng., Jan. 1960.
103. VOGE, J. Multiplex téléphonique et câbles hertziens, *An. Téléc.*, Feb. 1950.
104. WEISSMAN, L. see reference No. 94.
105. WILKINSON, E.J. Slot-antenna array for missiles and aircraft, *Electronics*, Feb. 1959.
106. WRIGHT, J.L. see reference No. 78.
107. ZELBSTEIN, V. *Applications industrielles des mesures électroniques*, Dunod, 1950.
108. Théorie du signal et de l'information et application aux télécommunications, *Bull. Ass. Suisse Electr.*, No. 12, 1953.

INDEX

Absorption, ionospheric V.3
Acceleration IV.3.3; IV.3.3.1; IV.3.3.2
Accelerometer, employing potentiometer VIII.2.5
Aerial V; V.5
  for a spinning missile V.5.3.3
  slot V.5.3.2
  spiral V.5.3.1
AJAX VIII.5.3
Angular aperture of an aerial V.1.1;
Area V.5
  effective, of an aerial V.1.1
  equivalent, for a radar echo V.6
Attenuation of propagation V.1.1

Bandwidth
  AM II.1.2
  FM II.2.1.2
  minimum VI.1.3
  PCM II.4.1
  PCM–AM II.4.1.2
  PCM–FM II.4.1.1
  pulse modulation II.3.2
Bessel function II.2.1.2
Boltzmann's constant IV.1.1.1

Capacitor, parallel plate VIII.2.4
Capacity of a transmission channel III.3
Capsule
  barometer IV.3.3.3
  with a membrane VIII.2.4
Circuits
  comparison III.8.1
  logic: "AND" II.4.4
  regenerative II.4.4
Climb, automatic VII.2.1
Code
  binary digits or bits III.7.2
  group III.7.3
  natural binary III.7.2
  reflected binary III.7.2
  sequence III.7.3
  Shannon–Fano III.7.3
Coder
  AN–AKT/14 VIII.5.2
  multiple frequency III.8.1
  with stages in cascade III.8.1
Coding
  analogue digital III.8.1
  digital III.7.2
  disc III.8.1
  matched III.3
  physical III.7.1
  secret III.7.4
  statistical III.7.3
Coding and decoding methods III.8
Coefficient
  band spreading III.4
  damping IV.3.3.2
  density II.3.1
  jamming IV.2.3
  reflection IV.4.2.2
  reflection of aerial feeder IV.4.2.3
  reflection of the ionosphere V.3
  transmission IV.3.3.2
Coincidence VI.2.2
Commutator VIII.4.3
Comparison of modulation VI.4
Control by mark–space variation VII.3.3
Converters VIII.4
  PAM–PDM II.3.3.2
  parallel–series code III.8.2
  series–parallel code III.8.3
  voltage frequency VIII.4.2
Cores with a rectangular hysteresis loop III.8.2
Criteria in telemetry VIII.1.3
Cross-modulation VI.2.1
  LF VI.2.1.1
  trailing VI.2.1.2
  various VI.2.1.3
Curie point IV.3.3.3

Decoding
  digital–analogue III.8.4
  of remote-control commands VII.3
Decommutation VIII.4.4
Decoupling of aerials V.5.2

Delay line VI.2.3
Density of information III.2
Detection
  AM II.1.4
  SSB II.1.5.1
  with correlation III.6.3
Detector, crystal II.1.4
Discriminator
  Foster–Seeley II.2.3.2
  Travis II.2.3.2
Dissipation in transistors IV.3.3.3
Distortion IV.4
  amplitude IV.4.1
  diversity V.7
  phase IV.4.2
  produced by a discriminator IV.4.1
  of a PCM signal in a filter II.4.3
Distribution or allocation of command signals VII.2.2
Diversity
  combiner V.7
  links with V.7

Echo reinforcer V.6
Efficiency
  information III.4; III.5
  signal-to-noise ratio
    SSB IV.1.4.1
    FM IV.1.4.2
    PAM IV.1.4.3
    PAM–FM IV.1.4.3
Electrical image V.5.2
E.m.f. of fluctuation IV.1.1.1
Entropy III.2
Errors I.4
  acceleration IV.3.3.2
  calibration IV.3.2
  dynamic IV.3.3
  hysteresis VIII.2.3
  mean IV.3.2
  mean square IV.3.2; VII.5.1
  of measurement IV.3
  quantization III.7.2; IV.3.2
  static IV.3.2
  temperature IV.3.3.3

Field propagation V
Filtering III.6.3
  optimum VII.5
Flapping ailerons VII.3.2; VII.3.3
Frequency
  instantaneous II.2.1.1
  natural VIII.2.5
Frequency (*cont.*)
  repetition II.3.1
  resonance IV.3.3.2
  sampling VI.2

Gain
  of an aerial V.1.1
  of a dipole V.4
  of the signal-to-noise ratio VI.1.4
Gaussian pulse II.3.2; II.4.1; II; Appendix
General remarks on missiles I
Gradient of the refractive index V.2.2
Group phase delay IV.4.2; IV.4.2.1; IV.4.2.2
Guidance I.2; VII
  mixed VII.5.3
Gyroscope, rate VIII.2.5

Hooke's Law VIII.2.1

Index
  AM modulation II.1.1.1
  amplitude distortion IV.4.1
  error IV.3.1
  FM modulation II.2.1.1
  modulation II.2.1.1
  refractive IV.1.2.1
Influence of external structure on polar diagram V.5.2
Information
  nature of III.1
  rate III.2; III.3
Interference IV.2.2
Intermodulation VI.1.1
  product VI.1.1
Ionosphere V.3

Jamming IV.2.3
Jet deflectors VII.3.3
Johnson effect IV.1.1.1

Limitation
  on weight and bulk I.5
  on information III.6.1; IV
Limiters
  amplitude II.2.3.1
  double diode II.2.3.1
  grid current II.2.3.1

Links, long distance III.6
Luneberg sphere V.6.2

Manometer VIII.2.3
Mark–space ratio VII.3.3
Missiles
  different types I.1
  spinning I.1
Modulation
  amplitude II.1
  anode or collector II.1.3.2
  by pulses II.3
  by varicap VII.4.3
  FM by a rectangular signal II.2.1.4
  frequency and phase II.2; VII.4.3
  grid or base II.1.3.1
  of pulses II.3.3
  PCM II.4
  PDM II.3.3.2
  PPM II.3.3.3
  SSB II.1.5
Modulator
  Armstrong II.2.2.3
  balanced II.1.5.1
  diode II.3.3.1
  reactance valves, transistors II.2.2.2
  serasoid II.2.2.3; VII.4.2
  valve (PAM) II.3.3.1
Multiplexing III.4; VI
  with distribution in frequency VI.1
  with distribution in time VI.2
Multiplexer VIII.4.3
Multivibrator, modulated VIII.4.2

Navigation
  positional VII.1
  velocity VII.1
Noise IV.1
  atmospheric IV.1.2.1
  cosmic IV.1.2.2
  environmental IV.1.2
  factor IV.1.1.4
  pulse IV.1.3
  semiconductor IV.1.1.3
  solar IV.1.2.2
Non-linearity in transducers VIII.3
Number system III.7.2
  binary III.7.2
  reasons for choosing binary III; Appendix

Optimization VII.5.1.
Order of diversity V.7
Oscillator VIII.4.1; VIII.4.2

Paraboloid aerial V.1.1
Pilot, automatic VII.1; VII.2.1
Piloting VII.2
Poisson distribution IV.1.1.2
Polarization V.2.1
Power
  minimum at receiver VI.3.1; VI.3.2
  of an AM wave II.1.1.2
  of an FM wave II.2.1.3
  of cosmic noise IV.1.2.2
Probe, altimeter VII.1
Production
  of AM waves II.1.3
  of FM and PM waves II.2.2
Propagation V
  equation V.3
  tropospheric V.2.2
  velocity IV.4.2.2
Pulse shaping VIII.4.4
Pulse width choice VI.2.1.4

Quantity
  measured by telemetry VIII.1.2
  of information III.2
Quantization of a signal III.3

Radiation diagram V.1.1; V.5
Radius, apparent, of the earth V.1.1
Range, maximum of link VI.3.3
Receiver, PCM II.4.4
Recorders VIII.6.1
Recording, magnetic VIII.6.2
Redundancy III.2; III.6.2
Reference magnitude IV.3.1
Reflection from the ground V.2.1
Reflector
  bistatic V.6.1
  monostatic V.6.1
Reluctance VIII.2.3
Remote-control VII
  FM–FM system VII.4
  receiver VII.4.4
  with commutation VII.3.3

Sampling VI.2.3
Selection of the channels VI.1.2

Self-optimization VII.5.2
Sensitivity factor of a strain-gauge VIII.2.1
Servomotor, height control VII.3.1
Shocks VI.3.3
Shot effect IV.1.1.2
Signal-to-noise ratio IV.1.3
  AM IV.1.4
  FM IV.1.4.2
  PAM, PAM–FM IV.1.4.3
  PCM IV.1.4.5
  PDM, PPM IV.1.4.4
  SSB IV.1.4.1
Skin temperature of missiles IV.3.3.3
Slicing VI.2.2
Spectrum
  AM II.1.2
  FM II.2.1.2
  of a train of pulses II.3.1
  PAM II.3.1
  PDM II.3.1
  PPM II.3.1
Spoiler VII.3.3
Stabilization
  against pitching VII.3.2
  against roll VII.3.2
  on course VII.3.2
Standing wave ratio V.5.1
Stores VIII.4.4
Strain-gauge VIII.2.1
Symbol III.2
Synchronization II.4.2; VI.2.4

Target missile VII.3.2
Telemetry equipment VIII
  PCM–FM VIII.5.2
  standards VIII; Appendix
Threshold range of link VI.3
  in measurements IV.3
  of limiter II.2.3.1
Teaching VII.1
Transducers VIII.2
  photoelectric VIII.2.2.2
  piezoelectric VIII.2.2.1
  potentiometers VIII.2.5
  with variable capacitance VIII.2.4
  with variable inductance VIII.2.3
Transmission channel III.1
Transmission, PCM II.4.3
Transmitter, PDM–FM telemetry VIII.5.1
Trigger, Schmitt II.4.4
Trirectangular trihedron V.6.1
Troposphere V.2.2
Tuller–Shannon theorem III.3

Uncertainty in a physical quantity IV.3.1

Variable capacitance transducer VIII.2.4
Varicap IV.3.3.3

Weighting III.7.2
Wobulation of a jammer IV.2.3

Copyrighted Material

Development of the Blue Streak Satellite Launcher

Edited by D. R. Samson

Department of Mechanical and Aeronautical Engineering, Hatfield College of Technology, Hertfordshire.

This book contains the details of the papers read at the second Space Engineering Symposium organized by the Department of Mechanical and Aeronautical Engineering at Hatfield College of Technology, Hertfordshire.
The book comprises six papers. The subjects covered include: performance; autopilot design; instrumentation; structural design; engine and propellant systems and the development of a launching site. It is to be noted that the statements and views contained here are those of the authors only, and should not be taken as the official views of the Ministry of Aviation or the European Launcher Development Organization.
Because of the extent to which work on the subject of satellites has been carried out in the United States of America and the Soviet Union, it is not always appreciated that Britain too, has been engaged in considerable research and has amassed a notable amount of knowledge about design and construction techniques. This book will provide a greater understanding of the contribution which Britain has made, and is making, to the important subject of space research.
International Series of Monographs in Aeronautics and Astronautics, Symposia Division Volume 16

"*While of first interest to the specialist, the average technical layman will find plenty to understand. A very good reference for military engineers.*"
Royal Engineers Journal.
60*s. net*/$9.00

Pergamon Press

Headington Hill Hall, Oxford
44–01 21st Street, Long Island City, New York 11101.
4 & 5 Fitzroy Square, London W1
2 & 3 Teviot Place, Edinburgh 1
6 Adelaide Street East, Toronto
24 rue des Ecoles, Paris 5e
Vieweg & Sohn GmbH, Burgplatz 1, Braunschweig, F.D.R.

The Flight of Uncontrolled Rockets

F. R. Gantmaker and L. M. Levin

A systematic course on the external ballistics of uncontrolled rockets, with consideration also given to the trajectory of shells, is presented in this volume. Emphasis is placed on the powered section of the trajectory, since conventional artillery methods are applicable to the remainder of the flight, and particular attention is paid to the treatment of rocket dispersion. The basic trajectory is calculated by treating the rocket as a particle subject to thrust, gravity and aerodynamic forces; the question of dispersion is then treated by considering the perturbations about the basic trajectory due to thrust malalignment, wind and other forces.
Finned non-rotated, finned rotated, and spin-stabilized rockets are dealt with in detail and attention is given to the special case of anti-tank rockets (short range finned rockets).
Russian methods differ in some respects from those used in the West, and for this reason a chapter on British and American work has been added by Mr. E. T. J. Davies which compares the methods and presents some additional results, while the first chapter has been completely revised and rewritten by Dr. A. J. Sarnecki, in order to make it specifically suitable for Western scientists.
The book forms a valuable first course in the subject, as it gives a very clear delineation of the fundamental principles involved. It should be on the shelves of libraries in universities, research establishments and industries in this important and specialized field.
International Series of Monographs in Aeronautics and Astronautics, Division VII Astronautics, Volume 5

£5 *net*/$15.00

"*. . . the book can be recommended as a sound introductory text . . . the general layout of the text maintains the usual high standards of the Pergamon Press and the price is reasonable.*"
Spaceflight.

Copyrighted Material